



Dual RNA-seq of pathogen and host

Duale RNA-Sequenzierung eines Pathogens und seines Wirts

Dissertation

zur Erlangung des naturwissenschaftlichen Doktorgrades (Dr. rer. nat.)

der *Graduate School of Life Sciences (GSLS)*,

Bayerische Julius-Maximilians-Universität Würzburg,

Sektion: Infektion und Immunität

vorgelegt von

Alexander J. Westermann

aus Bruchsal

Würzburg, im Dezember 2014



Eingereicht am: 8.12.2014

Mitglieder der Promotionskommission:

Vorsitzender: Prof. Dr. Thomas Dandekar

1. Gutachter: Prof. Dr. Jörg Vogel

2. Gutachter: Prof. Dr. Thomas Hünig

3. Gutachter: Prof. Dr. Thomas Rudel

Tag des Promotionskolloquiums:

Doktorurkunden ausgehändigt am:

Summary

The infection of a eukaryotic host cell by a bacterial pathogen is one of the most intimate examples of cross-kingdom interactions in biology. Infection processes are highly relevant from both a basic research as well as a clinical point of view. Sophisticated mechanisms have evolved in the pathogen to manipulate the host response and *vice versa* host cells have developed a wide range of anti-microbial defense strategies to combat bacterial invasion and clear infections. However, it is this diversity and complexity that makes infection research so challenging to technically address as common approaches have either been optimized for bacterial or eukaryotic organisms. Instead, methods are required that are able to deal with the often dramatic discrepancy between host and pathogen with respect to various cellular properties and processes. One class of cellular macromolecules that exemplify this host-pathogen heterogeneity is given by their transcriptomes: Bacterial transcripts differ from their eukaryotic counterparts in many aspects that involve both quantitative and qualitative traits. The entity of RNA transcripts present in a cell is of paramount interest as it reflects the cell's physiological state under the given condition. Genome-wide transcriptomic techniques such as RNA-seq have therefore been used for single-organism analyses for several years, but their applicability has been limited for infection studies.

The present work describes the establishment of a novel transcriptomic approach for infection biology which we have termed "Dual RNA-seq". Using this technology, it was intended to shed light particularly on the contribution of non-protein-encoding transcripts to virulence, as these classes have mostly evaded previous infection studies due to the lack of suitable methods. The performance of Dual RNA-seq was evaluated in an *in vitro* infection model based on the important facultative intracellular pathogen *Salmonella enterica* serovar Typhimurium and different human cell lines. Dual RNA-seq was found to be capable of capturing all major bacterial and human transcript classes and proved reproducible. During the course of these experiments, a previously largely uncharacterized bacterial small non-coding RNA (sRNA), referred to as STnc440, was identified as one of the most strongly induced genes in intracellular *Salmonella*. Interestingly, while inhibition of STnc440 expression has been previously shown to cause a virulence defect in different animal models of Salmonellosis, the underlying molecular mechanisms have remained obscure. Here, classical genetics, transcriptomics and biochemical assays proposed a complex model of *Salmonella* gene expression control that is orchestrated by this sRNA. In particular, STnc440 was found to be involved in the regulation of multiple bacterial target mRNAs by direct base pair interaction with consequences for *Salmonella* virulence and

implications for the host's immune response. These findings exemplify the scope of Dual RNA-seq for the identification and characterization of novel bacterial virulence factors during host infection.

Zusammenfassung

Die Infektion einer eukaryontischen Wirtszelle mit einem bakteriellen Pathogen ist eines der komplexesten Beispiele einer Domänen-überschreitenden Wechselwirkung zweier Organismen. Infektionsprozesse sind in höchstem Maße relevant, sowohl in der Grundforschung als auch von einem klinischen Blickwinkel aus betrachtet. Im Laufe der Evolution entstanden komplizierte Mechanismen, die es einem Pathogen erlauben, die Wirtsantwort zu manipulieren. Umgekehrt haben potentielle Wirtszellen eine Reihe von anti-mikrobiellen Verteidigungsstrategien entwickelt, um bakterielle Infektionen zu bekämpfen und letztlich zu beseitigen. Es sind jedoch genau diese Verschiedenheit und Komplexität, welche die Infektionsforschung so anspruchsvoll und technisch schwer analysierbar machen. Gängige Analysemethoden wurden zumeist entweder für bakterielle oder aber eukaryontische Organismen entwickelt. Dagegen werden Techniken benötigt, welche es erlauben, mit den mitunter extremen Unterschieden zwischen Wirt und Pathogen umzugehen, die sich in etlichen zellulären Eigenschaften und Prozessen manifestieren. Eine Klasse zellulärer Makromoleküle, die diese Heterogenität zwischen Wirt und Pathogen widerspiegelt, sind ihre jeweiligen Transkriptome: Bakterielle Transkripte unterscheiden sich von ihren eukaryontischen Pendants in vielerlei Hinsicht, was sowohl quantitative als auch qualitative Aspekte miteinschließt. Die Gesamtheit zellulärer Transkripte ist von größter Bedeutung, da sie den physiologischen Zustand der jeweiligen Zelle unter den gegebenen Bedingungen reflektiert. Aus diesem Grund werden Genom-weite Transkriptom-techniken wie etwa die RNA-Sequenzierung seit etlichen Jahren erfolgreich angewandt, um biologische Prozesse zu untersuchen – jedoch ist deren Eignung für Infektionsstudien in starkem Maße limitiert.

Die vorliegende Arbeit beschreibt die Etablierung eines neuartigen Ansatzes, „Duale RNA-Sequenzierung“ genannt, der Transkriptomstudien mit der Infektionsbiologie kompatibel macht. Mithilfe dieser Technologie wurde hier im Besonderen versucht, die Rolle nicht-proteinkodierender RNA-Moleküle für die Virulenz zu beleuchten, da die Charakterisierung dieser RNA-Klassen bisherigen Infektionsstudien weitgehend verwehrt blieb. Die Anwendbarkeit der Dualen RNA-Sequenzierung wurde innerhalb eines *In-vitro*-Infektionsmodells getestet, welches auf dem wichtigen, fakultativ intrazellulären Pathogen *Salmonella enterica* serovar Tyhimurium und verschiedenen humanen Zelllinien basiert. Die Duale RNA-Sequenzierung zeigte sich dabei in der Lage alle wesentlichen bakteriellen sowie humanen Transkriptklassen zu erfassen und erwies sich als reproduzierbar. Im Zuge dieser Experimente wurde ein Gen für eine zuvor kaum beschriebene kleine nicht-kodierende RNA (STnc440) als eines der am stärksten induzierten Gene intrazellulärer Salmonellen identifiziert. Interessanterweise hatten vorherige

Studien gezeigt, dass die Inaktivierung dieses Gens zu einem Virulenzdefizit innerhalb unterschiedlicher Tiermodelle für Salmonellose führt. Die zugrunde liegenden molekularen Mechanismen blieben jedoch unbekannt. In der vorliegenden Arbeit wurden genetische, Transkriptom- sowie biochemische Analysen eingesetzt um das komplexe Regulationsnetzwerk dieser kleinen RNA erstmals näher zu beleuchten. Im Einzelnen konnte gezeigt werden, dass STnc440 die Expression mehrerer bakterieller mRNAs durch das Ausbilden zwischenmolekularer Basenpaarungen reguliert, was weitreichende Konsequenzen sowohl für die Virulenz des Pathogens als auch die Immunantwort des Wirts hat. Diese Ergebnisse veranschaulichen das Potential der Dualen RNA-Sequenzierung für das Auffinden und Charakterisieren neuer bakterieller Virulenzfaktoren während der Wirtsinfektion.

List of figures

1.1	Fundamental differences between probe-dependent and probe-independent approaches to gene expression analysis.	11
1.2	Current model for the regulation of SPI-1 and SPI-2.	17
1.3	Schematic of host-pathogen interplay based on the model system of mammalian epithelial cells infected by <i>S. Typhimurium</i>	18
1.4	A new approach to host-pathogen transcriptomics.	20
1.5	Estimation of the minimal sequencing depth requirements for Dual RNA-seq.	22
2.1	Experimental workflow established for the simultaneous transcriptome analysis of <i>Salmonella</i> -infected mammalian cells.	25
2.2	Fluorescence microscopy of HeLa-S3 infected with GFP-expressing <i>S. Typhimurium</i>	26
2.3	Measuring bacterial dose effects on host cell infectivity, apoptosis and cytotoxicity.	27
2.4	Tracking proliferation of intracellular <i>S. Typhimurium</i> using diverse methods.	28
2.5	Evaluation of different transcriptome fixation techniques.	30
2.6	Global profile of RNA _{later} -fixed transcriptomes highly resembles those from unfixed cells.	31
2.7	GFP-based separation of infected from non-infected HeLa-S3 cells by cell sorting results in a high enrichment of the sub-populations.	33
2.8	Estimation of the fraction of bacterial RNA in the mixed host-pathogen transcriptome samples.	34
2.9	Computational work-flow for Dual RNA-seq.	35
2.10	Minimal material requirements for Dual RNA-seq.	37
2.11	Experimental overview of pilot Dual RNA-seq.	39
2.12	Spike-in RNA standards are used as an internal quality measure of Dual RNA-seq. ..	40
2.13	Read length distribution and cross-mapping reads.	41
2.14	Overview of Dual RNA-seq mapping statistics.	42
2.15	RPKM distribution.	43
2.16	Reproducibility of Dual RNA-seq.	44
2.17	Influence of the medium shift on <i>Salmonella</i> gene expression.	45
2.18	Pathogen dose effects on infectivity and cell death of human THP-1.	46
2.19	<i>Salmonella</i> virulence gene expression patterns in different host cell types.	47
2.20	Expression of genes from pathogenicity islands SPI-3 to -5.	48
2.21	A switch in <i>lpp</i> expression upon HeLa-S3 invasion.	48
2.22	Metal ion scarcity and membrane stress get reflected in the transcriptome of intracellular <i>Salmonella</i>	49
2.23	A high bacterial cell density is accompanied by initiator tRNA cleavage.	51
2.24	Intracellular sRNA expression profile.	53
2.25	Overlap in <i>Salmonella</i> gene expression between the infection of epithelial and monocytic cells.	54
2.26	Overlap of differentially regulated human mRNAs between invaded and bystander cells.	56
2.27	Immune genes are strongly induced after invasive infection.	56
2.28	NFκB targets and mitochondrial genes are induced in invaded HeLa-S3 cells.	58

2.29	Differential expression of human lncRNAs after <i>Salmonella</i> infection.	59
3.1	The STnc440 locus.	62
3.2	The STnc440 locus contributes to virulence in pig and cattle models of Salmonellosis.	63
3.3	Mild effect of STnc440 deletion on intracellular replication within HeLa-S3 cells.	64
3.4	Overview of the high-resolution Dual RNA-seq time-course of HeLa-S3 infection.	65
3.5	Intracellular sRNA expression profile of wild-type <i>S. Typhimurium</i> during the time-course of infection of HeLa-S3 cells.	66
3.6	STnc440 expression is strongly induced during infection.	66
3.7	STnc440 is strongly induced after the invasion of diverse human and mouse cell types.	67
3.8	<i>Salmonella</i> pathway analysis for the comparative RNA-seq experiment between wild-type and Δ STnc440 mutants <i>in vitro</i>	69
3.9	Influence of STnc440 on SPI-2 expression <i>in vivo</i>	70
3.10	STnc440 affects the accumulation of the SPI-2 effector protein SteC <i>in vitro</i>	70
3.11	<i>In vivo</i> sRNA pulse-expression establishment.	71
3.12	Pulse-expression experiments to identify target genes of STnc440.	72
3.13	Western blot analysis for STnc440 targets.	73
3.14	Direct targeting of <i>grxA</i> and <i>crp</i> mRNAs by STnc440.	74
3.15	STnc440 counter-acts SPI-2 activation through CRP <i>in vitro</i>	75
3.16	Infection with Δ STnc440 <i>Salmonella</i> impacts on <i>SOCS3</i> induction of the host.	76
3.17	STnc440 affects host STAT3 phosphorylation.	77
3.18	STnc440-dependent changes in host cytokine/chemokine levels.	77
4.1	Preliminary working model for the role of STnc440 in host-pathogen interplay.	83
7.1	<i>In vitro</i> validation of intracellularly induced sRNAs.	128
7.2	sRNA expression profile inside various host cell types.	129

List of tables

1.1	Properties and functions of RNA classes that constitute the bacterial and mammalian transcriptome.	4
2.1	cDNA libraries analyzed in the Dual RNA-seq pilot experiment.	39
3.1	cDNA libraries generated for the temporally resolved Dual RNA-seq experiment.	65
5.1	List of instruments and devices used.	90
5.2	List of lab ware used.	91
5.3	List of chemicals, reagents and proteins used.	92
5.4	Commercial kits used.	93
5.5	List of enzymes used.	94
5.6	List of primary and secondary antibodies and -sera used.	94
5.7	List of synthetic oligonucleotides used.	95
5.8	List of plasmids used.	97
5.9	List of <i>Salmonella</i> strains used.	97

Abbreviation index

(NH ₄) ₂ SO ₄	ammonium sulfate
(v/v)	(volume/volume)
(w/v)	(weight/volume)
A	adenine
a.k.a.	also known as
AGE	agarose gel electrophoresis
Amp	ampicillin
AP-1	activator protein-1
APS	ammonium persulfate
asRNA	anti-sense RNA
ATP	adenosine triphosphate
BSA	bovine serum albumin
Cm	chloramphenicol
CDS	coding sequence
C	cytosine
c.f.u.	colony-forming units
CAGE	cap analysis of gene expression
cDNA	complementary DNA
CIP	calf intestinal phosphatase
circRNA	circular RNA
CLIP	cross-linking and immunoprecipitation
co-IP	co-immunoprecipitation
CRP	cyclic AMP receptor protein
crRNA	CRISPR-associated RNA
DMEM	Dulbecco's Modified Eagle Medium
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dNTP	deoxyribonucleotide triphosphate
dRNA-seq	differential RNA sequencing
ds	double-stranded
DTT	dithiothreitol
<i>E. coli</i>	<i>Escherichia coli</i>
EcnB	entericidin B
EDTA	ethylenediamine tetraacetic acid
EGFP	enhanced green fluorescent protein
EGTA	ethylene glycol tetraacetic acid
ELISA	enzyme-linked immunosorbent assay
endo-siRNA	endogenous small interfering RNA
EtOH	ethanol
Fig.	figure
FCS	fetal calf serum
G	guanine

gDNA	genomic DNA
GFP	green fluorescent protein
GO	gene ontology
grxA	glutaredoxin 1
GTP	guanosine triphosphate
Hfq	host factor of bacteriophage Q β
hnRNP	heterogeneous nuclear ribonucleoprotein
IGB	Integrated Genome Browser
IGR	intergenic region
IL	interleukin
JAK	Janus kinase
Kan	kanamycin
KEGG	Kyoto Encyclopedia of Genes and Genomes
LB	Lennox broth
LDH	lactate dehydrogenase
lincRNA	long intergenic RNA
lncRNA	long non-coding RNA
LPS	lipopolysaccharide
M	million
m.o.i.	multiplicity of infection
MACE	massive amplification of cDNA ends
MAPK	mitogen-activated protein kinase
mRNA	messenger RNA
miRNA	microRNA
n.a.	not applicable
n.d.	not detected
NaCl	sodium chloride
NAT	natural anti-sense transcript
NCBI	National Center for Biotechnology Information
NF κ B	nuclear factor-kappa B
OD	optical density
p	phosphorylated
P:C:I	phenol:chloroform:isoamyl alcohol
p.i.	post-infection
PAA	polyacrylamide
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PFA	paraformaldehyde
PI3K	phosphatidyl-inositol-3-kinase
piRNA	PIWI-interacting RNA
PMA	phorbol 12-myristate 13-acetate
PNPase	polynucleotide phosphorylase
qRT-PCR	quantitative real-time PCR
RBP	RNA-binding protein
RBS	ribosome binding site
RISC	RNA-induced silencing complex
RNA	ribonucleic acid

RNase	ribonuclease
RNA-seq	RNA sequencing
RNP	ribonucleoprotein complex
ROS	reactive oxygen species
rRNA	ribosomal RNA
<i>S. Typhi</i>	<i>Salmonella enterica</i> serovar Typhi
<i>S. Typhimurium</i>	<i>Salmonella enterica</i> serovar Typhimurium
SAGE	serial analysis of gene expression
scRNA	small cytoplasmic RNA
scaRNA	small Cajal body RNA
SCV	<i>Salmonella</i> -containing vacuole
SD	standard deviation
Ser	serine
siRNA	small interfering RNA
snoRNA	small nucleolar RNA
snRNA	small nuclear RNA
SOC	Super Optimal broth with Catabolite repression
SOCS	Suppressor of cytokine signaling
SPI	<i>Salmonella</i> pathogenicity island
SPIA	Signal Pathway Impact Analysis
sRNA	small RNA
ss	single-stranded
SSC	saline-sodium citrate
STAT	Signal transducer and activator of transcription
STnc	<i>Salmonella</i> Typhimurium non-coding RNA
T	thymine
T3SS	type III secretion system
TA	toxin/anti-toxin
TAP	tobacco acid phosphatase
TBE	Tris/Borate/EDTA
TEMED	tetramethylethylenediamine
TraDIS	transposon directed insertion-site sequencing
Tris	tris-(hydroxymethyl)-aminomethane
tRNA	transfer RNA
trxA	thioredoxin 1
TSS	transcription start site
tracrRNA	trans-activating crRNA
Tyr	tyrosine
U	uracil
UTR	untranslated region
vol.	volume

Units

%	percent
°C	degree Celsius
A	Ampère
bp	base pair
Da	Dalton
g	gravity
g	gram
h	hour
L	liter
M	molar
min	minute
nt	nucleotide
RPKM	Reads per kilobase transcript per million mapped reads
rpm	rounds per minute
sec	second
U	unit
V	Volt
W	Watt

Dimensions

f	Femto (10^{-15})
p	Pico (10^{-12})
n	Nano (10^{-9})
μ	Micro (10^{-6})
m	Milli (10^{-3})
k	Kilo (10^3)
M	Mega (10^6)
G	Giga (10^9)

Table of contents

Summary	I
Zusammenfassung	III
List of figures.....	V
List of tables	VII
Abbreviation index	VIII
Units	XI
Dimensions	XII
Table of contents	XIII
1. Introduction	1
1.1 The transcriptome as a snapshot of an infected cell's physiological state	1
1.2 Transcriptomics: then and now	8
1.3 RNA-seq: variation of a theme	11
1.4 Towards a joint analysis of pathogen and host transcriptomes	15
1.5 <i>Salmonella</i> as a model organism for host-pathogen interaction studies.....	16
1.6 Dual RNA-seq: a thought experiment.....	19
1.7 Aim of the study	23
2. Implementation of the Dual RNA-seq approach.....	24
2.1 Choice of the infection model and optimization of infection conditions.....	24
2.2 Transcriptome fixation	29
2.3 FACS-based enrichment of infected cells.....	32
2.4 Computational challenges.....	34
2.5 Defining the minimal material requirements for a Dual RNA-seq experiment.....	36
2.6 A proof-of-principle experiment	38
2.7 <i>Salmonella</i> gene expression changes revealed by Dual RNA-seq.....	44
2.8 Changes in the host transcriptome	55
2.9 Concluding remarks	60
3. The STnc440 sRNA regulates <i>Salmonella</i> virulence gene expression with impact on the host response.....	61
3.1 STnc440 inhibition coincides with a fitness defect in different infection models.....	61
3.2 STnc440 expression is rapidly induced after invasion of various host cell types.....	64
3.3 Regulation of <i>Salmonella</i> virulence genes by STnc440.....	68
3.4 JAK/STAT and cytokine signaling are STnc440-affected host pathways.....	75
3.9 Concluding remarks	78
4. Discussion	79
4.1 Dual RNA-seq: potential and scope for future improvements.....	79
4.2 Rewiring the regulatory network of virulence gene expression in <i>Salmonella</i>	81
4.3 The STnc440 regulon may modulate the host response towards an appeased state.....	84
4.4 The concept of “molecular phenotypes” of bacterial sRNAs	85
4.5 Future applications.....	86
4.6 Conclusions and perspectives	89
5. Material and methods	90
5.1 Material.....	90
5.2 Methods.....	101

6. References	114
7. Appendix	114
7.1 Intracellular sRNA expression profiles in diverse host cell types	128
8. Curriculum vitae	128
9. List of publications	131
10. Attended conferences and courses	132
11. Contribution by others	135
12. Acknowledgements	135
13. Affidavit/Eidesstattliche Erklärung	137

1. Introduction

1.1 The transcriptome as a snapshot of an infected cell's physiological state

Currently, several possibilities exist to globally analyze biological processes within bacterial or eukaryotic cells and these can be grouped into four fundamentally differing approaches: (1) Genomics addresses the genetic material (i.e. the genomic DNA) of an organism, while (2) transcriptomics focusses on the transcribed proportion of the genome (i.e. RNA) and (3) proteomics assesses the proteinaceous fraction of cells (i.e. what is translated from RNA into protein). Finally, (4) metabolomics detects cellular metabolites to describe the distinct biological processes of a cell or organism. All these approaches have been shown to yield invaluable information regarding the physiological state of cells or their responses to external perturbations and changing environments. During the past decade, however, it became increasingly clear that the traditional dogma of gene expression – which assigns effector functions only to proteins – is obsolete. Instead diverse classes of RNA transcripts are now known to fulfill important functions that go far beyond serving as simple messengers for protein synthesis or as house-keeping molecules. In bacteria diverse non-coding RNA elements have been identified, including anti-sense RNAs (Thomason and Storz, 2010), riboswitches (Winkler and Breaker, 2005) or small RNAs (sRNAs; see below). In higher eukaryotes, as demonstrated by the Encyclopedia of DNA elements (ENCODE) project, protein-coding genes constitute only a very minor fraction (~3%) of the genome and are outnumbered by non-coding genes (Dunham et al., 2012). Together, this collective body of work has provided a compendium of non-coding transcripts (see Tab. 1.1) of which the most relevant classes for the present study will be introduced in detail in the following sections.

1.1.1 Bacterial sRNAs

Bacteria possess diverse non-coding RNAs (see Tab. 1.1). Among them the class of sRNAs constitutes the largest and best-studied cohort, with representatives in all studied bacterial species (both Gram-negative and -positive) (Gottesman and Storz, 2011; Storz et al., 2011). Many sRNA genes show signs of horizontal transfer, but conservation studies are generally difficult as sequence constraints are limited and often the secondary structure defines functionality (Gottesman and Storz, 2011). In *Salmonella*, the most recent transcriptome-wide analysis has predicted the presence of close to 300 different sRNAs (Kroger et al., 2013). sRNA biogenesis is heterogenic but generally it can be broadly classified into two mechanisms: Classically, sRNA genes are transcribed independently of the neighboring genes from their own

promoters. Others, however, have recently been reported to be processed from the untranslated regions (UTRs) of coding genes (Chao et al., 2012).

sRNAs are defined as short transcripts (generally between 50 and 500 nt) that lack coding potential. Rather, these transcripts are well known to have effector functions themselves, e.g. by post-transcriptionally regulating (in most known cases repressing) the expression of target mRNAs in *cis* or *trans*. Typically regulation is mediated by base-pairing between the so-called seed region of the sRNA and critical sites in the mRNA target – in the classical scenario the ribosome binding site, thereby preventing translation initiation and inducing mRNA decay. For this to occur, many sRNAs rely on a chaperone protein, referred to as Hfq (host factor of the RNA bacteriophage Q β) (Vogel and Luisi, 2011). Hfq facilitates inter-molecular RNA-RNA interactions, recruits specific ribonucleases for target mRNA degradation and protects the bound sRNA from being nucleolytically attacked, thereby increasing its stability (Wagner, 2013). Accordingly, this sub-class of transcripts has been termed Hfq-dependent sRNAs. Recently, the RNA chaperone ProQ has emerged as a second global sRNA binder, with Hfq and ProQ dependence not being mutually exclusive (Smirnov et al., unpublished). Finally, instead of direct base-pairing with target mRNAs, other members of the class of sRNAs have been demonstrated to function as titrators by binding and sequestering proteinaceous regulators of gene expression. The best-studied representatives of this sub-class are two sRNAs referred to as CsrB and CsrC, which both bind to the translational regulator CsrA (Carbon storage regulator A), thereby modulating the activity of this protein (Babitzke and Romeo, 2007; Romeo et al., 2013).

Irrespective of their precise mode of action, sRNAs seem to generally fine-tune and/or time target gene expression, rather than mediating strict on/off switches as typically done by transcription factors. sRNAs often help bacteria to rapidly adapt to changing environmental cues and they have been implicated in cellular responses to diverse stress conditions (Benjamin et al., 2010; Bobrovskyy and Vanderpool, 2014; Repoila et al., 2003; Vogel and Papenfort, 2006). Consequently, they often act within complex regulatory circuits (Hoe et al., 2013; Vogel and Papenfort, 2006) including virulence programs (see section 1.1.4.).

1.1.2 Eukaryotic microRNAs (miRNAs)

Initially discovered in worms, miRNAs are meanwhile known to be present in a wide range of eukaryotic organisms, from flies and plants to mammals. In humans, there are currently close to 2,000 mature miRNAs annotated (miRBase; as of October 2014). Dependent on the organism, miRNAs are about 21-25 nt in length. They are typically transcribed as long, polyadenylated pri^{mary} transcripts (pri-miRNAs) and processed in the nucleus into pre^{cursor} stem-loops (pre-miRNAs), which are subsequently exported into the cytosol (Krol et al., 2010). Here further processing occurs by the RNase III-type endonuclease Dicer and results in a double-stranded

(ds) miRNA duplex. The mature guide strand is eventually loaded together with a member of the Argonaut protein family into the RNA-induced silencing complex (RISC). As part of RISC, the guide miRNA scans transcripts and recruits the complex to the respective target mRNAs to cleave and degrade, translationally inhibit and/or deadenylate them (He and Hannon, 2004). Thus, several differences notwithstanding, by acting in concert with defined protein chaperones and ribonucleases as post-transcriptional regulators of mRNA translation and stability, eukaryotic miRNAs appear functionally analogous to bacterial sRNAs.

1.1.3 Eukaryotic long non-coding RNAs (lncRNAs)

Recently, lncRNAs have emerged as an additional eukaryotic non-coding transcript class with regulatory capacity. Unlike miRNAs, lncRNAs constitute a heterogenic class of molecules. By definition, all transcripts >200 nt that lack coding potential are referred to as lncRNAs. Current predictions for human transcriptomes indicate that there might be >10,000 individual representatives (Derrien et al., 2012). For practical reasons, the class of lncRNAs is therefore typically further sub-divided into long intergenic non-coding RNAs (lincRNAs) as well as sense and anti-sense lncRNAs that overlap with protein-coding loci. Individual candidates may or may not be polyadenylated or spliced and localize to the nucleus and/or the cytosol. Likewise, the reported modes of action for this transcript class differ substantially: An emerging theme in lncRNA-mediated control of gene expression, however, seems to be the recruitment of histone-modifying complexes to specific loci in the genome to epigenetically regulate target gene transcription (Mercer and Mattick, 2013; Rinn, 2014). For several further lncRNAs, however, other functions have been reported: Some representatives act as decoys by titrating transcription factors and thus, indirectly regulate target gene transcription (Hung et al., 2011). Others guide protein complexes to their correct sub-cellular destination (Huarte et al., 2010) or play structural roles by forming a scaffold for the assembly of cellular sub-compartments (Souquere et al., 2010). In some cases, not even the product but rather the act of transcription itself provides functionality: For instance, deep-sequencing has identified the existence of a class of divergent lncRNAs that are transcribed anti-sense to the promoters of coding genes and provide directionality of coding gene transcription (Almada et al., 2013). The resulting lncRNA products might simply represent byproducts of transcription and are typically rapidly degraded after synthesis (Clark et al., 2012).

Table 1.1/Properties and functions of RNA classes that constitute the bacterial and mammalian transcriptome.

	transcript class	approximate proportion of total RNA	size range	biogenesis	nature of 5'/3'end	major reported function(s)	Infection-relevant representative(s) ¹	reference(s)
bacteria	rRNA	~60-90% (depending on metabolic state)	120 nt (5S); ~1,500 nt (16S); ~3,000 nt (23S)	transcribed as longer precursor followed by extensive modifications, the cleavage and maturation of both ends	5'-mono-phosphate; 3'-hydroxyl	component of ribosomes (mRNA translation)	-	
	tRNA	~5-15%	~70 nt	transcribed as longer precursor followed by extensive modifications and the maturation of both ends; a specific amino acid is attached to each tRNA	5'-mono-phosphate; 3' CCA + hydroxyl group	anti-codon adapter molecule for mRNA translation	initiator tRNA cleavage by virulence-associated VapC protein	(Winther and Gerdes, 2011)
	mRNA	~5%	varies (100's to 10,000's of nt)	primary transcripts (mono- or polycistronic)	5'-triphosphate; 3'-hydroxyl	template for protein synthesis	-	
	tmRNA	~1%	~360 nt	transcribed as a larger precursor and processed into mature tmRNA (similar to tRNA biogenesis)	processed and aminoacylated (similar to tRNAs)	rescues and recycles stalled ribosomes; mRNA/protein quality control ("trans-translation")	-	(Himeno et al., 2014)
	6S RNA	~1%	~200 nt	transcribed from own promoter; further processed on either end	5'-mono-phosphate; 3'-hydroxyl	global regulator of transcription by interacting with the σ^{70} RNA polymerase holo-enzyme	<i>Legionella pneumophila</i> 6S RNA was shown to impact on intra-macrophage replication	(Cavanagh and Wassarman, 2014; Faucher et al., 2010)
	sRNA	~5%	~50-500 nt	either transcribed from own promoter or processed from UTRs of mRNAs	diverse 5' ends (tri- or monophosphorylated depend on biosynthesis as primary transcript or processing product); 3'-hydroxyl	gene regulation by direct (imperfect) base-pair interaction with mRNAs or titration of RNA-binding proteins	e.g. <i>S. Typhimurium</i> : IsrM, IsrJ, IsrI, IsrH-1, OxyS, InvR, DapZ, SgrS, RyhB, IsrE; <i>E. coli</i> : MgrR; <i>Staphylococcus aureus</i> : SprD; <i>Vibrio cholerae</i> : TarA/B, VrrA; <i>Listeria monocytogenes</i> : SreA; <i>Chlamydia trachomatis</i> : IhtA	(Albrecht et al., 2010, 2011; Chabelskaya et al., 2014; Chao et al., 2012; Gong et al., 2011; Grieshaber et al., 2006; Mann et al., 2012; Moon et al., 2013; Ortega et al., 2012; Padalon-Brauch et al., 2008; Pfeiffer et al., 2007)
	crRNA, (tracrRNA)	species-dependent (only some bacteria have a functional CRISPR/Cas system)	~24-48 nt (crRNAs); ~30-100 nt (tracrRNA)	crRNA synthesis very diverse and dependent on CRISPR type (e.g. processed from longer precursor or separately transcribed from intrinsic promoters); tracrRNA transcribed from own promoter	for crRNAs dependent on mode of biogenesis; tracrRNA with 5'-triphosphate, 3'-hydroxyl	genome defense; regulation of endogenous gene expression	CRISPR/Cas-mediated regulation of bacterial lipoprotein expression in <i>Francisella novicida</i>	(Bhaya et al., 2011; Heidrich and Vogel, 2013; Sampson et al., 2013; van der Oost et al., 2014)
	cis-asRNA	under debate (often depends on technological aspects but also species-specific differences)	from tens to thousands nt	independently transcribed	5'-triphosphate; 3'-hydroxyl	under debate (gene regulation by antisense mechanisms, "excludon" concept, role in genome evolution as asRNAs tend to be enriched in genomic regions of foreign origin, source of nanoRNAs?, transcriptional noise?)	e.g. <i>S. Typhimurium</i> AmgR, IsrC, lesR-1; <i>Shigella flexneri</i> RnaG	(Georg and Hess, 2011; Giangrossi et al., 2010; Gonzalo-Asensio et al., 2013; Gottesman and Storz, 2011; Lasa et al., 2011; Lee and Groisman, 2010; Padalon-Brauch et al., 2008; Sesto et al., 2013)
	circRNAs	rare	varies	arise from the 3'-5' ligation of both ends of linear RNA molecules	n. a.	-	-	(Doose et al., 2013; Vicens and Cech, 2009)
	nanoRNA	unknown	~2-4 nt	unknown	5'-hydroxyl	priming transcription initiation (incorporated into nascent mRNA)	-	(Nickels, 2012)
	bacterial lncRNAs	unknown	several hundred nt	independently transcribed	5'-triphosphate; 3'-hydroxyl	regulation of gene expression from opposite strand (excludon concept)? enzymatic activity?	-	(Conway et al., 2014; Weinberg et al., 2009)

Table 1.1/continued.

	transcript class	sub-class	approximate proportion of total RNA	sub-cellular localization	size range	biogenesis	nature of 5'/3'end	major reported function(s)	Infection-relevant representative(s) ¹	reference(s)
mammals	rRNA	n. a.	~60-90% (depending on metabolic state)	mitochondria + cytoplasm	121 nt (5S); 156 nt (5.8S); ~1,900 nt (18S); ~5,000 nt (28S); mitochondrial: ~1,000 nt (12S) and ~1,600 nt (16S)	pre-rRNAs transcribed by RNA Pol I; extensively processed and modified into mature rRNAs	5'-mono-phosphate; 3'-hydroxyl	component of ribosomes (mRNA translation)	-	
	tRNA	n. a.	~5-15%	mitochondria + cytoplasm	~70-90 nt	pre-tRNAs transcribed by RNA Pol III; extensively processed and modified into mature tRNAs; a specific amino acid is attached to each tRNA	5'-mono-phosphate; 3' CCA + hydroxyl group	anti-codon adapter molecule for (cytoplasmic) mRNA translation	tRNA cleavage products (Angiogenin-mediated processing) might be involved in gene regulation	(Li et al., 2012b)
	mRNA	n. a.	~5%	translated in cytoplasm; may be shuttled into P-bodies to pause translation or degrade mRNA	varies (10's to 100,000 nt)	RNA Pol II transcripts as pre-mRNAs; co-transcriptionally 5'-capped, spliced and 3'-polyadenylated	5'-capped (7-methylguanylate); 3'-poly(A) tail and -hydroxyl group	template for protein synthesis	-	
	snRNA	n. a.	together <5%	nucleus (within "splicing speckles")	~150 nt	transcribed by RNA Pol II as 5'-capped pre-snRNAs, exported into the cytoplasm for extensive modification e.g. hypermethylation of 5' cap (except for U6 snRNA which is transcribed by PolIII and does not enter the cytoplasm), nuclear re-import and further modifications (pseudouridylation, methylation)	5'-capped (2,2,7,-methylguanylate); 3'-hydroxyl	pre-mRNA processing/splicing	-	
	snoRNA	C/D box		nucleolus	~70–120 nt	typically spliced from intronic regions of coding transcripts that are transcribed via RNA Pol II and further processed (exonucleolytic trimming), but also independently transcribed representatives		rRNA maturation (guide methylation)	-	
		H/ACA box		nucleolus	~150-200 nt			rRNA maturation (guide pseudouridylation)	-	
	scaRNA	n. a.		Cajal bodies (nuclear organelle)	~150-200 nt			guide modifications (methylation and pseudouridylation) of spliceosomal RNAs	-	
	scRNA	n. a.	~1-2%	cytosol	diverse, but generally <300 nt	diverse	diverse	diverse; typically associate with cytoplasmic proteins to form RNP complexes involved in regulation of the last step of gene expression	-	

Table 1.1/continued.

transcript class	sub-class	approximate proportion of total RNA	sub-cellular localization	size range	biogenesis	nature of 5'/3'end	major reported function(s)	Infection-relevant representative(s) ¹	reference(s)
miRNA	n. a.	typically ~0.1-2% (in exceptions up to 10%)	cytosol, nucleus, mitochondria	21-22 nt	transcribed as a primary transcript (pri-miRNA) in the nucleus (in exceptions, spliced out of a coding transcript: "Mirtron"), cleaved into pre-miRNA, exported into the cytosol, cleaved into miRNA duplex and mature guide miRNA loaded into RISC	5'-mono-phosphate; 3'-hydroxyl	gene regulation by direct (imperfect) base-pairing with RNA transcripts	miR-146, miR-155, let-7, miR-15, miR-21, miR-371-3, miR-29, miR-125a/b, miR-149	(Eulalio et al., 2012)
endo-siRNA	n. a.	unknown	nucleus	<30 nt	endogenous dsRNA substrates are chopped into smaller fragments	5'-mono-phosphate; 3'-terminally modified	silence transposable elements in germ and somatic cells; role in heterochromatin formation; much better characterized in plants than in mammals	-	(Nilsen, 2008)
piRNA	n. a.	cell type-specific	nucleus	~26-30 nt	primary piRNA as a product of long single stranded precursor molecules; secondary piRNAs generated via "Ping Pong" mechanism	5'-mono-phosphate; 2'-O-methyl modification at the 3' end	genome defense (transposon silencing) in the germ line	-	(Aravin et al., 2007)
l(i)ncRNA	poly(A) ⁺	~5% (but precise fraction unclear)	nucleus, cytoplasm, cellular sub-compartments (P-bodies, paraspeckles, etc.)	extremely heterogenic (<0.5 kb to >100 kb)	independently transcribed; might or might not be spliced and polyadenylated	5'-capped (7-methylguanylate); 3'-hydroxyl	diverse (gene regulation by interacting with chromosomal DNA or RNA transcripts, components of RNP complexes, structural components of sub-cellular compartments, etc.)	NeST, Lethe, lincRNA-Cox2, THRIL, NEAT1, IL1β-RBT46	(Carpenter et al., 2013; Gomez et al., 2013; Iotti et al., 2014; Imamura et al., 2014; Li et al., 2014; Rapicavoli et al., 2013)
	poly(A) ⁻							-	
cis-NATs	n. a.	under debate (often depends on technological aspects)	nucleus, cytoplasm	varies	typical RNA Pol II-dependent transcripts	5'-capped (7-methylguanylate); 3'-hydroxyl	transcriptional interference, RNAi	-	
circRNAs	n. a.	unknown	predominantly in the cytoplasm	varies	arise from the 3'-5' ligation of both ends of linear RNA molecules	n. a.	"sponge" for miRNAs	-	(Jeck and Sharpless, 2014; Memczak et al., 2013)

¹: in the given context of bacterial infection of eukaryotic host cells

crRNA: CRISPR-associated RNA

tracrRNA: trans-activating crRNA

asRNA: anti-sense RNA

circRNA: circular RNA

scaRNA: small Cajal body RNA

scRNA: small cytoplasmic RNA

endo-siRNA: endogenous small interfering RNA

piRNA: PIWI-associated RNA

NATs: natural anti-sense transcripts

n.a.: not applicable

1.1.4 Non-coding RNAs involved in virulence, infection and immunity.

Representatives from each of these classes of non-coding RNAs are emerging as important regulators in infection processes (Tab. 1.1): For instance, in bacterial pathogens a wealth of sRNAs have been found to play profound regulatory roles not only in physiology but also pathogenicity (reviewed in Caldelari et al., 2013; Gripenland et al., 2010; Hebrard et al., 2012; Papenfort and Vogel, 2010, 2014; Toledo-Arana et al., 2007). In *Salmonella*, genome-wide expression analyses under various infection-related conditions have begun to identify potential virulence-related sRNAs. For instance, the FUR-regulated sRNAs RyhB and IsrE have been found to be highly induced in a fibroblast model of *Salmonella* infection (Ortega et al., 2012). Many pathogenicity island-encoded sRNAs (*Isr* genes) are activated inside macrophages (Gong et al., 2011; Padalon-Brauch et al., 2008) and numerous additional sRNAs undergo strong expression changes during growth over a range of infection-mimicking *in vitro* conditions (Kroger et al., 2013). Moreover, *in vitro* studies with *Salmonella* and pathogenic *E. coli* suggest various sRNAs that contribute to pathogenicity. These include the *Salmonella* pathogenicity island 1 (SPI-1)-associated InvR (Pfeiffer et al., 2007) and DapZ (Chao et al., 2012) sRNAs, the lipopolysaccharide (LPS) synthesis-controlling MgrR (Moon et al., 2013), or the virulence effector-targeting SgrS sRNA (Papenfort et al., 2013). However, inhibition of many of these well characterized sRNA genes by transposon insertions did not lead to a pronounced growth defect in *in vivo* models of infection (Barquist et al., 2013; Chaudhuri et al., 2013). Rather, these global transposon mutagenesis screens suggest that several additional, so-far under-characterized sRNAs compromise *Salmonella* virulence when deleted – with one such example being *Salmonella* Typhimurium non-coding RNA 440 (STnc440).

Also in the eukaryotic cell many non-coding transcripts are known to fulfill regulatory functions, and this includes the orchestration of the immune response upon pathogenic attack: In various cell types representatives of the miRNA family have been shown to regulate among others the expression of key immune genes, cell-cycle genes or enzymes involved in oxidative stress (reviewed in Eulalio et al., 2012). For instance, the conserved miR-146 and miR-155, whose expression is induced upon Toll-like receptor 4 (TLR4) stimulation by bacterial LPS (Taganov et al., 2006), feedback to regulate TLR4 signaling in a non-redundant manner (Schulte et al., 2013). The let-7 miRNA family was shown to regulate cytokine gene expression after *Salmonella* infection of murine macrophages (Schulte et al., 2011). In addition, members of the miR-15 family have recently been reported to be suppressed in epithelial cells by invading *Salmonella*, thereby making host cells more susceptible to infection (Maudet et al., 2014).

Representatives of the heterogenic class of lncRNAs are now emerging as crucial regulators for a variety of cellular processes, with a few reported examples of immunity-associated lncRNAs.

Pioneering studies reported individual lncRNAs to regulate immunity processes by means of interacting with key transcription factors such as nuclear factor- κ B (NF κ B) (Rapicavoli et al., 2013) or heterogeneous nuclear ribonucleoproteins (hnRNPs) (Carpenter et al., 2013; Li et al., 2014). Recently, the application of global RNA-seq to LPS-stimulated monocytic cells was used to compile a catalogue of immune-relevant lncRNA candidates and revealed that in some cases the identified lncRNAs affect the expression of important immune genes in *cis* (Liott et al., 2014).

These studies notwithstanding, compared to their protein-encoding counterparts, non-coding transcripts are still under-characterized in the context of infection. The previous findings, however, demonstrate infection to be reflected not only in the coding transcriptome but also in the non-coding RNA expression profile of both a eukaryotic host cell and a pathogenic bacterium. Therefore, the steady-state composition of all cellular RNAs would represent an invaluable proxy for the physiological state of both a given bacterial and eukaryotic cell and thus, unbiased transcriptomic approaches promise unprecedented access to a global understanding of host-pathogen interplay.

1.2 Transcriptomics: then and now

Traditional methods to measure the cellular expression of individual transcripts were based on gene-specific DNA probes used for detection by hybridization (e.g. Northern blot) or by amplification (e.g. quantitative real-time PCR (qRT-PCR)). However, such approaches were limited to the analysis of a small number of pre-selected target genes. Global quantitative transcriptomics became possible upon the development of microarrays. These made use of numerous DNA probes that had been immobilized in parallel on glass slides or membranes (Fodor et al., 1993; Schena et al., 1995). Fluorescently labeled RNA (or corresponding cDNA) samples were subsequently hybridized to the microchip and after washing, the abundance of any given transcript in the input sample could be inferred from the fluorescent signal intensity of the spot corresponding to the respective gene probe. This led to a shift from hypothesis-driven to discovery-driven science. While microarrays allowed the study of expression states of multiple genes simultaneously, they were still limited to the detection of known RNA molecules (often mRNAs) and therefore failed to discover novel transcripts. This was in part overcome by the introduction of high-resolution tiling arrays (Bertone et al., 2004; Selinger et al., 2003). These relied on the same working principle as microarrays but instead of spotting DNA probes for defined transcripts, stretches of genomic DNA were immobilized on the chip surface, thereby not only allowing for the detection of known transcripts but also for the identification of novel gene products. However, despite their great potential, the high costs associated with tiling

arrays – especially when analyzing gene expression in organisms with large genomes such as higher eukaryotes – restricted their application to a few specialized laboratories. Additionally, since expression states are deduced from fluorescent signal intensities that have both a lower (sensitivity) and an upper limit (saturation), both microarrays and tiling arrays have a low dynamic range and are often restricted to the analysis of the most abundant cellular transcripts (typically mRNAs) (Wang et al., 2009).

Alternatives to probe-dependent methods that are occasionally employed for eukaryotic transcriptomes are tag-based techniques. The original approach, termed serial analysis of gene expression (SAGE) (Velculescu et al., 1995), combined 3'-primed reverse transcription of cellular RNAs with the identification of the resulting cDNA fragments ('tags') by Sanger sequencing. These short (13-15 nt) cDNA tags corresponding to a transcript's 3' end were then assigned to their respective source gene and enumerated to inform about its expression state. A modification of the technique to selectively sequence tags from the 5' cap of eukaryotic mRNAs referred to as cap analysis of gene expression (CAGE) (Kodzius et al., 2006) was introduced later in order to identify transcription start sites (TSSs) in eukaryotic genomes on a global scale. More recently, the problem associated with unequivocally assigning short tags to their specific locus in the genome has been partially overcome with the improvement of SAGE to SuperSAGE (Matsumura et al., 2003) and massive amplification of cDNA ends (MACE) (Asmann et al., 2009) which generate longer tags (25-94 nt) than the former approaches. Whereas tag-based approaches, by counting individual sequence tags, entail a much increased dynamic range as compared to fluorescence-dependent protocols, they still suffer from rather low sensitivity (e.g. they are unsuitable to profile a whole transcriptome including diverse non-coding transcripts). Noteworthy, all currently available tag-based approaches are restricted to the detection of either the 3' (SAGE, SuperSAGE, MACE) or 5' end (CAGE) of input RNA molecules, thereby preventing the full-length coverage of cellular transcripts and failing to discriminate between different transcript isoforms.

The above-described transcriptomic approaches have been widely applied to various experimental systems including infection models. For instance, the first genome-wide transcriptomic studies of several important bacterial pathogens such as *Vibrio cholerae* (Merrell et al., 2002), *Borrelia burgdorferi* (Revel et al., 2002), *Chlamydia trachomatis* (Belland et al., 2003), *C. pneumonia* (Maurer et al., 2007) and *S. Typhimurium* (Eriksson et al., 2003) were performed with the help of microarrays and revealed gene expression patterns that suggested strategies used by these microbes to adapt to their host cells. Likewise, tiling arrays were applied to both *in vitro* and *in vivo* grown *Listeria monocytogenes* (Toledo-Arana et al., 2009) and used to identify sRNAs or novel virulence genes in streptococci (Kumar et al., 2010; Perez et al., 2009; Zheng et al., 2011). In turn, microarrays yielded important insights into the global

response of potential mammalian hosts upon cytokine (Zhu et al., 1998) or pathogen-associated molecular pattern (PAMP) stimulation (Hedges et al., 2005; Kerns et al., 2009; Tross et al., 2009), as well as towards diverse bacterial (reviewed in Rappuoli, 2000) or viral infections (Der et al., 1998). Recently, two microarray-based studies reported the response of human epithelial cells to *S. Typhimurium* infection and found that by secreting effector proteins *Salmonella* induces pervasive transcriptional reprogramming in the host (Bruno et al., 2009; Hannemann et al., 2013). Less frequently, tag-based approaches have been applied to the field of infection biology, e.g. to profile gene expression in eukaryotic parasites (reviewed in Kronstad, 2006) or in group A *Streptococcus pyogenes*-infected human epithelial cells that, in combination with microarray analysis, revealed the induction of a set of apoptosis-related mRNAs in invaded host cells (Nakagawa et al., 2004).

In summary, even though these diverse transcriptomic approaches provided novel insights into the infection process of eukaryotic cells with diverse pathogens, they all suffer from the high amounts of input RNA required as well as the rather low sensitivity and/or dynamic range. Most importantly, the inherent characteristic of the described approaches to be mRNA-centric prevented the performance of “whole”-transcriptome studies (i.e. to profile both coding and non-coding gene expression in an unbiased fashion).

Next-generation sequencing (NGS) technologies have recently revolutionized biology. Illumina/Solexa, Roche/454 (a.k.a. Pyrosequencing) or ABI/SOLiD were the first NGS technologies to enter the market. Even though these systems have very different working principles they all share a dramatic increase in speed as compared to conventional Sanger-sequencing: Long DNA stretches or even entire genomes could be sequenced within several hours (Metzker, 2010). Initially applied to genomics, since ~2008 NGS technologies have been adapted for the massively parallel sequencing of cDNA, which is referred to as RNA-seq. The advent of RNA-seq enabled the high-throughput profiling of entire transcriptomes and has heralded a new era in the field of transcriptomics (reviewed in Wang et al., 2009). Compared to traditional approaches, the digital nature of RNA-seq (Fig. 1.1) makes it highly sensitive and provides a wide dynamic range. In addition, RNA-seq does not (necessarily) require prior annotation of the corresponding genome (even though it benefits from it), it is organism-independent and bears the potential to capture all (coding and non-coding) cellular transcripts. Importantly, not only can quantitative information about gene expression (i.e. differential expression) be obtained, but also qualitative information: RNA-seq allows for the refinement of transcript boundaries, the identification of UTRs, processing sites, splice isoforms in eukaryotes or bacterial operon structures, respectively, the screening for TSSs as well as the identification of novel (e.g. non-coding) genes. All these features have readily rendered RNA-seq the method of choice for transcriptomic studies irrespective of the organism of interest.

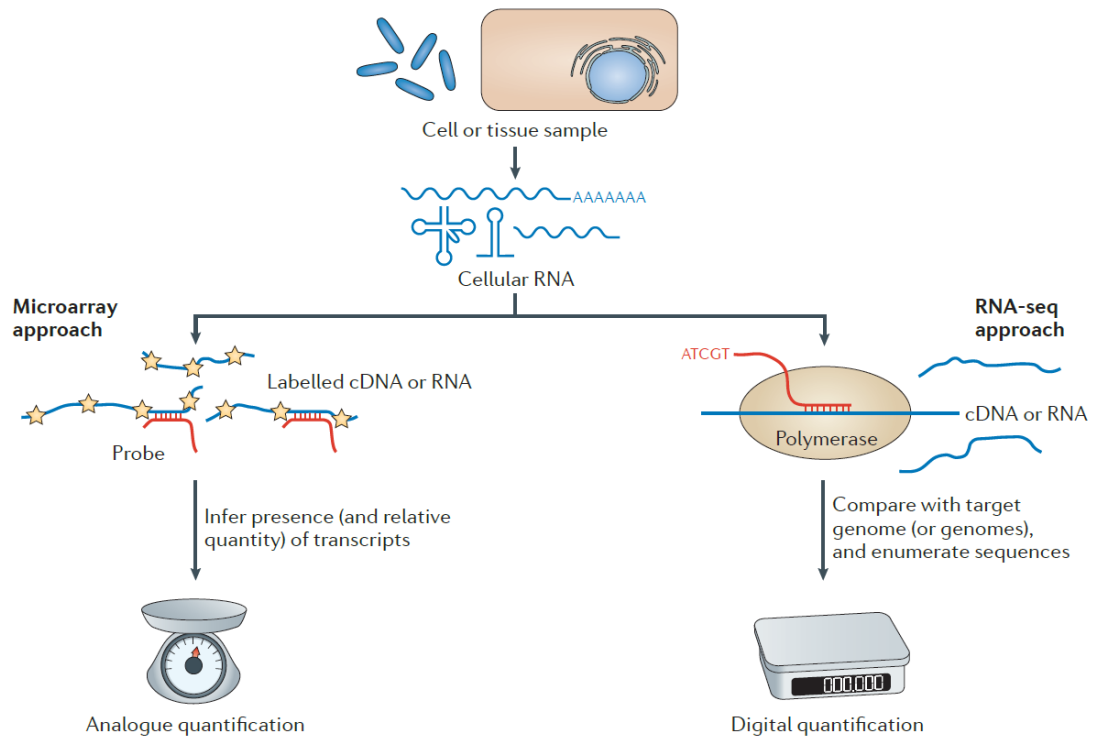


Figure 1.1|Fundamental differences between probe-dependent and probe-independent approaches to gene expression analysis. Probe-dependent methods such as microarrays rely on fluorophore labeling of RNA or cDNA, which allows for an analogue quantification of gene expression that is limited by a narrow dynamic range and a lower and upper detection threshold. In contrast, sequencing of cDNA molecules enables the digital enumeration of reads and increases both sensitivity and dynamic range. Figure taken from Westermann et al., 2012.

1.3 RNA-seq: variation of a theme

RNA-seq is based on the simple principle that by sequencing cDNA molecules on NGS platforms the expression state of any cellular RNA transcript can be inferred from the read count mapped to it. The “standard” RNA-seq protocol thus starts with extracting cellular RNA from a biological sample. Total RNA is next reverse transcribed into cDNA, amplified and then adapter or linker sequences are added to either end of each cDNA molecule to render it compatible with the respective sequencing technology. Sequencing of the final cDNA library results in the generation of “sequencing reads”, which are sequence stretches of a defined length (depending on the NGS technology they can vary between ~50 and a few hundreds of nucleotides) that correspond to the input cDNA. In a last step, sequencing reads are usually mapped to the reference genome sequence of the respective organism (if available). Thus, by enumerating the mapped reads per annotated gene its relative expression state under the sampled condition can be determined. Several additional steps have been introduced to improve the basic protocol and increase the information obtained from RNA-seq data. Some of the most important concepts will be outlined below.

RNA enrichment/depletion: Ribosomal transcripts usually comprise the majority of RNA molecules (up to 90%) in both bacterial and eukaryotic cells. However, due to their house-keeping role in cellular protein synthesis they are typically constitutively expressed and thus their expression state is of low informative value. Consequently, protocols were developed to either deplete rRNA or selectively enrich the mRNA fraction prior to sequencing, thereby minimizing sequence reads derived from unwanted transcripts and maximizing those from the desired RNA classes. Depletion/enrichment can either be achieved on the RNA level or – upon reverse transcription – on the cDNA level. For instance, the commercial RiboMinus, Ribo-Zero and MICROBExpress kits are based on DNA oligonucleotides with sequence complementarity to rRNA that were immobilized on magnetic beads for the efficient removal of ribosomal transcripts from total RNA samples (reviewed in Sultan et al., 2014). Likewise, enzymatic treatment (e.g. exonuclease digestion for bacterial transcriptomes) can be used to specifically deplete processed transcripts, including the abundant mature rRNA and tRNA molecules, while primary transcripts remain (Sharma et al., 2010). Alternatively, instead of their active removal rRNAs can passively be depleted during cDNA synthesis by the use of ‘not-so-random’ primers which represent a mixture of all possible but ribosomal sequences to prime reverse transcription (Armour et al., 2009).

Rather than depleting rRNA (and tRNA), eukaryotic transcriptomes can be enriched for polyadenylated mRNAs and lncRNAs. Thus, kits based on a similar principle to the described rRNA depletion technique but using oligo(dT)-coupled magnetic beads have been used to capture and enrich poly(A)⁺ RNAs. In contrast, studies investigating short RNA species (e.g. bacterial sRNAs or eukaryotic miRNAs) often involve a size-fractionation of total RNA on a gel or column to exclude longer RNA molecules (including the larger ribosomal transcripts) from reverse transcription.

All these techniques substantially lower the required sequencing depth (see below). On the other hand, the enzymes or kits are expensive, the required amounts of starting material is increased compared to total RNA-seq and it is often unclear if or to what extent technical biases are introduced during the additional steps (The Sequencing Quality Control Consortium, 2014). Consequently, several studies deliberately refrain from depletion/enrichment but sequence total RNA.

Fragmentation: Without prior treatment, RNA/cDNA molecules can usually not be covered over the entire transcript body using the common deep-sequencing technologies that are based on short reads. Thus, to ensure homogeneous coverage and to eliminate length biases during PCR amplification, nucleic acids may be fragmented prior to sequencing. Different fragmentation methods exist that shear either the original RNA or the corresponding cDNA molecules into

smaller, uniform fragments. Generally, these can be divided into three principle approaches: (1) to mechanically fragment the sample (e.g. via sonication), (2) to enzymatically cleave nucleic acids, or (3) to incubate them in a buffer that induces their chemical fragmentation. In practice, the deep-sequencing technology used determines the best-suited fragmentation strategy of a sample. However a general observation seems to be that transcript coverage is more homogeneous when fragmentation is performed on the RNA level, whereas cDNA shearing introduces a bias towards a transcript's 3' end (Wang et al., 2009).

Strand-specificity: In the earliest RNA-seq protocols information on the strand-specificity of the original RNA molecules was lost as a consequence of cDNA amplification. Upon first strand synthesis, the single-stranded (ss) cDNA had to be amplified to match the input amount requirements and thus was used as a template for PCR. This resulted in ds cDNA products that were then sequenced. The obtained sequencing reads could be mapped to a specific locus in the genome, but not assigned to one of the two strands. This is problematic because prokaryotic genomes are typically very densely staggered with genes on both the leading and lagging strand which frequently overlap one another. Likewise, anti-sense transcription is widespread in animals (Dunham et al., 2012). Therefore, current protocols rely on different techniques to maintain strand-specificity during cDNA library generation and sequencing (reviewed in Levin et al., 2010). Of these, the most common protocols involve the selective labeling of second strand cDNA that allows for its specific degradation after the library amplification. Alternatively, adapters may be ligated to the input RNA molecules to maintain information about their orientation during cDNA synthesis and amplification.

Depth requirements: The first RNA-seq studies in mammals were performed on the polyadenylated fraction of cellular transcripts and the resulting libraries were sequenced to ~10-50 M reads (Mortazavi et al., 2008; Sultan et al., 2008). The first studies in prokaryotes reported the sequencing of rRNA-depleted samples to ~1-7 M reads/library (Oliver et al., 2009; Perkins et al., 2009; Yoder-Himes et al., 2009). This was sufficient to infer differential mRNA expression. However, an increased depth is required for RNA-seq to provide information on the expression levels of low abundant RNAs, transcript boundaries, splice isoforms or operon structures. In fact, to deduce both quantitative and qualitative aspects of gene expression by RNA-seq, depth recommendations for an exhaustive profiling of bacterial or mammalian rRNA-depleted transcriptomes suggest 1-10 M or ~100 M reads/library, respectively (Haas et al., 2013; Toung et al., 2011). Obviously, when refraining from depletion/enrichment of individual RNA classes but sequencing total RNA instead, a greater depth will be required to reach the same coverage.

In 2008 the first studies reported RNA-seq-based transcriptomics for murine (Mortazavi et al., 2008) and human cells (Sultan et al., 2008), and thus for potential host organisms of bacterial infections. These early studies focused on coding gene expression as reverse transcription was primed using oligo(dT) primers to selectively sequence only polyadenylated transcripts (mRNAs or polyadenylated lncRNAs). These works uncovered the sheer complexity of mammalian genome expression, including pervasiveness of transcription, a multitude of lncRNAs as well as the prevalence of alternative splicing and isoform variation. Meanwhile, RNA-seq has been coupled to diverse upstream methods to give rise to tailored techniques for answering specific scientific questions: For instance, combining it with the pull-down of RNA-binding proteins by means of co-immunoprecipitation (co-IP) or *in vivo* cross-linking and immunoprecipitation (CLIP) allows for the global detection of transcript binding partners of cellular RNA-binding proteins (RBPs) or the identification of novel RBPs (e.g. Castello et al., 2012). A modification of this approach to specifically pull-down nascent ribosomes or RNA polymerase molecules provides a snapshot of translating mRNAs (Ingolia et al., 2009) or nascent transcripts (Churchman and Weissman, 2011), respectively, at a given time point in the cell. Finally, even the secondary structure of cellular transcripts can be investigated with the help of deep-sequencing, namely by analyzing RNA samples that were treated with different structure-specific enzymes (Kertesz et al., 2010).

RNA-seq technology was subsequently adapted to bacterial transcriptomes, essentially by circumventing the need for polyadenylated transcripts as substrates for reverse transcription and by decreasing RNA input material requirements. Many of the pioneering studies were performed using pathogenic bacteria, including *S. Typhi* (Perkins et al., 2009), *Bacillus anthracis* (Passalacqua et al., 2009), *Burkholderia cenocepacia* (Yoder-Himes et al., 2009), *Helicobacter pylori* (Sharma et al., 2010) or *Chlamydia trachomatis* (Albrecht et al., 2010), and identified a plethora of novel genes including many new sRNAs (for reviews see (Croucher and Thomson, 2010; Guell et al., 2011; Sorek and Cossart, 2010). The original RNA-seq protocol has since been further modified to increase the informational output: For instance, the usage of a specific exonuclease in a method referred to as differential RNA-seq (dRNA-seq) allowed the selective degradation of processed RNA molecules leaving only primary transcripts, thereby enabling the identification of TSSs on a genome-wide scale (Sharma et al., 2010). The transcript interactome of specific RNA-binding proteins could be analyzed on global scale by combining co-IP with RNA-seq. For instance, studies with *S. Typhimurium* reported the sRNA interactome of the important bacterial RNA chaperone Hfq over growth *in vitro* (Chao et al., 2012; Sittka et al., 2008). Collectively, these works have changed our concept of bacterial transcriptomes which turn out to be much more complex and flexible than previously anticipated. For example, many genes have more than just a single TSS and often external stimuli dictate which TSS will be used. Also

operon compositions are highly diverse, i.e. a given gene might be transcribed both as a monocistron and as part of a larger operon.

1.4 Towards a joint analysis of pathogen and host transcriptomes

During infection, bacterial pathogens are in intimate contact with their eukaryotic host cell. Thus, in order to be close to their natural niche, pioneering microarray studies aimed to analyze infection within joint host-pathogen models rather than in pure isolation (reviewed in Rappuoli, 2000). These studies however, came with their own intrinsic problems: In order to prevent cross-hybridization, chips were either designed to exclude probes with a high risk of detecting transcripts from both organisms or the data derived from such probes were later excluded from the data analysis. Cross-hybridization was particularly an issue when analyzing pathogen gene expression as the bacterial RNA is typically overwhelmed by the host transcriptome background. Consequently, protocols were developed to physically separate host and pathogen from one another and to enrich for bacterial RNA after infection had taken place but prior to detection (reviewed in Hinton et al., 2004). These enrichment techniques, which are still commonly used nowadays, either rely on the separation of bacterial from eukaryotic cells or transcripts. In the former case, if the model system is based on extracellular pathogens, the host-pathogen interaction is disrupted and differential centrifugation results in the selective sedimentation of host cells while the bacterial cells remain in the supernatant fraction. In case of invading pathogens, host cells can be selectively lysed (e.g. by detergent treatment). In such a scenario, differential centrifugation will pellet intact bacterial cells while the host lysate remains in the supernatant. Preparing RNA from the pellet or supernatant fraction, respectively, will result in an enrichment of pathogen RNA. Alternatively, discrimination between host and pathogen can be achieved on the transcript level. Methods such as the MICROBEnrich kit make use of the different characteristics of eukaryotic and bacterial RNA species. That is, similar to the rRNA removal method described above, oligonucleotides that are specific to the most abundant eukaryotic transcripts (18S rRNA, 28S rRNA, polyadenylated mRNAs) might be coupled to magnetic beads to capture and pull out a substantial fraction of host RNA from a mixed transcriptome sample. Finally, depletion of host material may be achieved on cDNA level via the selective capture of pathogen-derived cDNAs by means of hybridization using biotinylated DNA probes derived from the respective bacterial chromosome (Bent et al., 2013; Morrow et al., 1999).

In the context of their natural niche, transcriptomic analyses unraveled pathogen gene expression patterns that differed from bacteria grown *in vitro* (Mandlik et al., 2011). Likewise the host response to live bacterial challenge was found to be more complex than that towards

individual PAMPs (e.g. Hannemann et al., 2013). However, separation and enrichment are labor-intensive, error-prone as additional steps in the protocol are considered potential sources of bias, and usually are incompatible with low amounts of starting material. In contrast the high resolution of RNA-seq, down to the single-nucleotide level, achieves a high discriminatory power. In theory, this renders RNA-seq the ideal method to study host and pathogen gene expression without the need to physically separate the cells.

1.5 *Salmonella* as a model organism for host-pathogen interaction studies

Salmonella represents a genus of Gram-negative enterobacteria that comprises two species (*S. enterica* and *S. bongori*) with more than 2,500 individual serovars. The complete ~5 Mb genome sequence from various *Salmonella* strains is available (Parkhill et al., 2001). *Salmonella* spp. typically infect their hosts by the oral ingestion of contaminated food or drinking water. *Salmonella* are facultative intracellular pathogens and several serovars exhibit a relevant disease potential for humans. For instance, infection with primate-specific *S. enterica* serovar Typhi (*S. Typhi*) or *S. Paratyphi* can lead to lethal typhoid fever while *S. Typhimurium* has a broader host range including poultry and cattle, and infected humans may suffer from non-typhoid, self-limiting gastroenteritis. In the mouse model, however, *S. Typhimurium* causes systemic infections, thereby mimicking human typhoidal fever upon *S. Typhi* infection. Due to the ease of cultivation under standard laboratory conditions and its genetic similarity, but lower risk potential for human health as compared to other serovars, *S. Typhimurium* is amongst the best-studied Gram-negative model organisms. This includes diverse genetic and biochemical tools which are routinely applicable for the study of this organism.

Salmonella pathogenicity is mediated by effector proteins that are directly secreted via a specialized multi-component apparatus referred to as the type-three secretion system (T3SS) into the host's cytoplasm. Upon translocation, these effectors substantially manipulate important host cellular processes to the bacterium's benefit, including cytoskeletal rearrangements, membrane trafficking and signal transduction pathways (reviewed in Haraga et al., 2008). Two distinct T3SSs were acquired by *Salmonella* and are known to be essential for infection and host colonization: The T3SS of SPI-1 (T3SS-1) translocates bacterial effectors across the plasma membrane of the target cell to mediate invasion upon host contact. A complex regulatory circuit ensures precise expression of T3SS-1 by integrating various environmental parameters of the small intestine into the regulation cascade that converges on the level of the AraC-like transcriptional activator HilD (Ellermeier and Schlauch, 2007) (Fig. 1.2). Upon its stimulation, HilD transcriptionally activates two further regulators, HilC and RtsA, which amplify the signal, leading to the rapid induction of the entire SPI-1 locus.

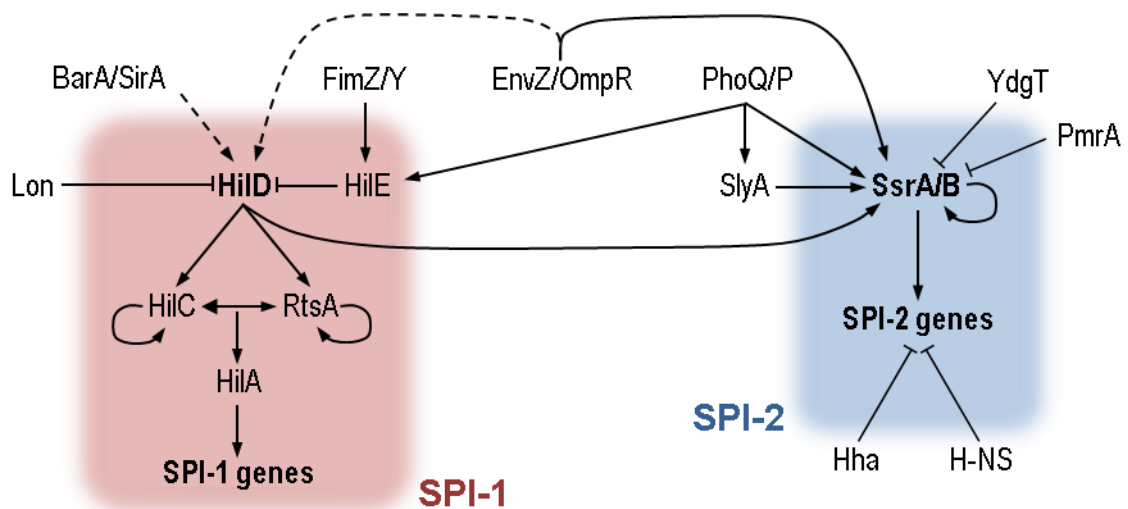


Figure 1.2|Current model for the regulation of SPI-1 and SPI-2. Regulatory network compiled from Choi and Groisman, 2013, Ellermeier and Slauch, 2007, and Fass and Groisman, 2009. Solid lines represent direct regulations and dashed lines represent indirect effects. See main text for details.

In contrast, the T3SS-2 routes effectors across the vacuolar membrane of *Salmonella*'s intracellular replication niche, the so-called *Salmonella*-containing vacuole (SCV). Thus, expression of SPI-2 has to be tightly controlled to ensure its activation only within the host cell environment. This precision is mediated by various upstream positive and negative regulators, including the two-component systems OmpR/EnvZ and PhoQ/PhoP, that sense environmental factors and feed into the regulatory network (Fass and Groisman, 2009) (Fig. 1.2). A third two-component system, SsrA/B, acts as a hub by integrating the activating or inhibiting stimuli and upon activation the response regulator protein SsrB binds directly to the promoters of SPI-2 genes to activate their transcription.

The two pathogenicity islands, however, do not operate independently (Fig. 1.2). For example, HilD, the master activator of SPI-1 genes has been shown to regulate transcription of the *ssrA/B* operon, encoding the key regulator of SPI-2 (Bustamante et al., 2008; Martinez et al., 2014). Conversely, the PhoP/Q system that activates SPI-2 expression is known to shut down SPI-1 activity (Bajaj et al., 1996; Behlau and Miller, 1993). These and similar mechanisms are considered to coordinate the cross-talk between the different SPIs, with huge impact on *Salmonella* virulence (Fabrega and Vila, 2013).

Conversely, mammalian cells possess diverse anti-microbial defense strategies to antagonize bacterial invasion which are based on the innate and adaptive branch of immunity (reviewed in Abdullah and Knolle, 2014; de Jong et al., 2012). Conserved PAMP motifs on the surface of pathogens are sensed by extra- or intracellular pattern recognition receptors (PRRs) of the host. Paradigm examples in this context represent the detection of bacterial LPS by TLR4, that of

flagellin by TLR5 or that of bacterial peptidoglycans by NOD-like receptors (NLRs). Consequently various intracellular pathogens, including *Salmonella*, have evolved strategies to minimize sensing by host PRRs, e.g. by down-regulating flagellin expression (Winter et al., 2008) or modifying LPS (Guo et al., 1997; Matsuura et al., 2012) after host cell invasion. Triggered by PAMP detection, both non-professional (i.e. epithelial cells) and professional immune cells (macrophages) undergo defined gene expression programs orchestrated by key transcription factors of immunity, such as activator protein-1 (AP-1) or NFκB. This may lead to the secretion of cytokines to activate defense programs in neighboring cells, to recruit phago- and granulocytes to the site of infection or to activate the adaptive immune response and generate immunological memory. Evolved as a host defense mechanism to clear infection, inflammation can be manipulated by certain pathogenic species including *Salmonella* that have evolved mechanisms to dampen or augment the host's inflammatory response. That is, in contrast to the host induction of an immune response by PRR-mediated PAMP sensing, these pathogens translocate effectors into their host cell to actively elicit a classical response. The pathogen may profit from the resulting environmental alterations, e.g. by inducing release of nutrients not present in non-inflamed tissues. For instance, effector proteins of *S. Typhimurium* can induce gut tissue inflammation, allowing the pathogen to tap host nutrients and outcompete members of the commensal microbiota that often lack the capability to utilize these compounds (Thiennimitr et al., 2012).

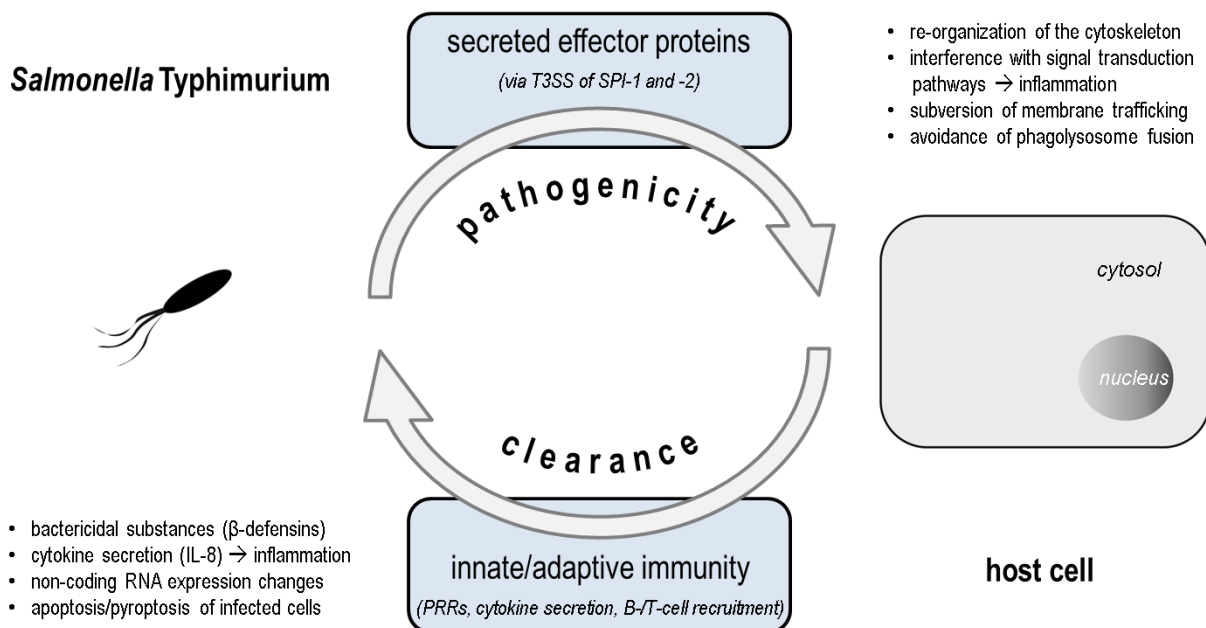


Figure 1.3/Schematic of host-pathogen interplay based on the model system of mammalian epithelial cells infected by *S. Typhimurium*. Inflammation was considered both a host defense mechanism as well as a strategy employed by the pathogen for successful colonization. See main text for further details.

S. Typhimurium appears to be an ideal choice as a model organism for the establishment of a novel host-pathogen transcriptomic approach. A wealth of previous data outlines a paradigm example for the adaptivity of cross-kingdom interaction processes (Fig. 1.3). As infection involves a tight interplay between two active opponents many aspects of the underlying biological processes may only be fully appreciated when analyzing the two partners jointly. This would potentially allow for the identification of novel host-pathogen interaction nodes and for a better understanding of the behavioral repertoire of the two counter-players during infection.

1.6 Dual RNA-seq: a thought experiment

A single-step protocol for the simultaneous profiling of host and pathogen transcriptomes is very attractive for several reasons: Omitting the physical separation of host and pathogen cells or of their endogenous transcripts would substantially speed up sample preparation (Fig. 1.4). Likewise, the fewer steps in the experimental workflow from sample acquisition to analysis would minimize the sources of technical bias. Most importantly, however, profiling pathogen gene expression within its host cell niche, together with the corresponding host response without disrupting this tight interaction would provide an unprecedented and unperturbed view of the infection process. Directly relating gene expression changes in one organism to responses in the respective counter-player might reveal novel mechanisms of how these organisms interact. For such an experiment, that would allow for the simultaneous analysis of two unrelated transcriptomes, we coined the term “Dual RNA-seq” (Westermann et al., 2012).

Transcriptomic heterogeneity anti-correlates with the relative ease of performing multi-organism RNA-seq. Such heterogeneity nowhere is more extreme than in the context of bacterial infections of mammalian host cells. It manifests itself in both quantitative and qualitative traits: A prototypical mammalian cell contains in the range of 10-20 pg of total RNA. This is >100 fold more when compared to a prototypical prokaryotic cell which harbors ~100 fg of RNA. In practical terms, this means that in order to capture gene expression changes in both organisms, an optimal proportion of bacterial to host cells should be present in the sample. Ideally both genomes should eventually be covered to a similar degree, in order to prevent from the hyper-profiling of one and the hypo-profiling of the other partner. A bacterial genome may be around 1-5 Mb in size, while that of humans contains around 3,000 Mb. Thus, to reach an equal coverage of both genomes, human reads should theoretically be in ~1,000 fold excess over the bacterial reads. However, bacterial genomes are usually densely packed with genes and – unlike their eukaryotic counterparts – there is little non-transcribed space, which makes the discrepancy less dramatic.

For infections with extracellular pathogens, the infectious dose should be adjusted in a way that allows for the parallel investigation of both infection partners. In the case of intracellular bacterial species, selecting for invaded and against non-infected bystander cells would probably facilitate the adjustment of a suitable host-to-pathogen proportion in the sample. This would further differentiate between the specific immune responses of infected host cells and non-infected – yet extracellularly stimulated – bystanders.

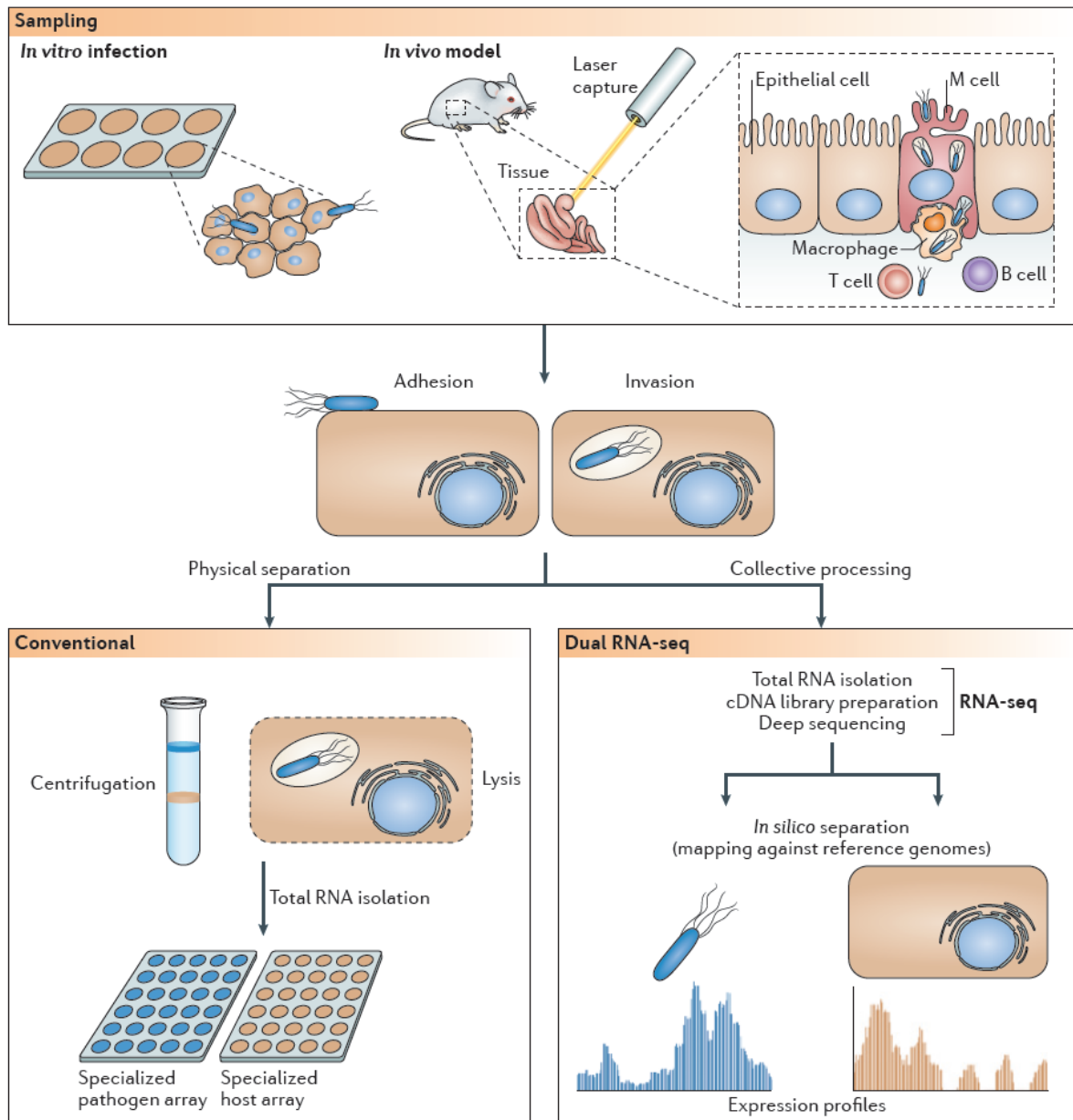


Figure 1.4|A new approach to host-pathogen transcriptomics. Traditional approaches for host pathogen transcriptomics (left) depended on the physical separation of host and pathogen prior to their analysis. Instead, Dual RNA-seq omits a separation step and analyzes the two organisms together (right). Discrimination between host and pathogen gene expression thus only occurs *in silico*, by mapping RNA-seq reads to the respective reference genomes. Figure taken from Westermann et al., 2012.

Quantitative differences are accompanied by qualitative heterogeneity between bacterial and mammalian transcriptomes. Even though house-keeping transcripts (rRNA, tRNA) are common in bacteria and eukaryotes, their specific properties may differ. Importantly, mRNAs show fundamentally different characteristics between the two kingdoms: Whereas bacterial mRNAs harbor a 5' triphosphate group, the 5' end of eukaryotic messengers is protected by a cap structure (7-methylguanosine). Additionally, although polyadenylation of messengers exists in bacteria and eukaryotes, the average number of adenines (A's) per mRNA molecule and importantly the fate of the corresponding transcript differ substantially (Dreyfus and Regnier, 2002): In eukaryotes, with the exception of histone mRNAs all messengers are co-transcriptionally polyadenylated. Mammalian mRNAs contain ~250 A's at their 3' end and this poly(A)-tail mediates stability of the transcript. In stark contrast, bacterial mRNAs are not routinely polyadenylated, but if they are, the poly(A)-stretch is short (~50 A's) and marks the respective transcript for degradation. Consequently, selection of the polyadenylated fraction would isolate transcripts with contrasting fates in the different organisms and therefore, enrichment for specific RNA classes is not recommended for Dual RNA-seq. Transcriptome heterogeneity further increases as there are several RNA classes that are exclusive to either bacterial or mammalian transcriptomes (Tab. 1.1). Furthermore, whereas polycistrons are common in bacteria, eukaryotic mRNAs are expressed as monocistrons. Splicing, RNA processing and editing or sub-cellular localization further add to the qualitative heterogeneity between bacterial and eukaryotic transcriptomes.

The intra- and inter-organism transcriptome heterogeneity notwithstanding, the conceptual idea of simultaneously profiling host and pathogen gene expression would be relevant for bacterial/mammalian infection models. Yet additional considerations would have to be taken for the experimental protocol to capture gene expression in both counter-players. For example, to account for the transcriptomic heterogeneity described above, one would refrain from the enrichment and/or depletion of certain RNA classes, but sequence total RNA instead. The RNA extraction protocol should efficiently isolate both bacterial and eukaryotic transcripts, at best irrespective of their size. During cDNA library construction it is essential to include a fragmentation step in order to shear the diverse RNA molecules into more homogenous fragments. Importantly, strand-specificity should be guaranteed to be able to assess expression of anti-sense transcripts in both organisms. Moreover, as an internal quality control, biological samples should be supplemented with artificially synthesized RNA molecules (spike-in RNAs) (Jiang et al., 2011). The primary sequence of any such spike-in RNA must be confirmed bioinformatically to be absent from all the genomes under investigation. Features such as spike-in RNA length, GC content or concentration have to be empirically optimized for any assay system. Lastly, the required sequencing depth should be estimated for each infection model.

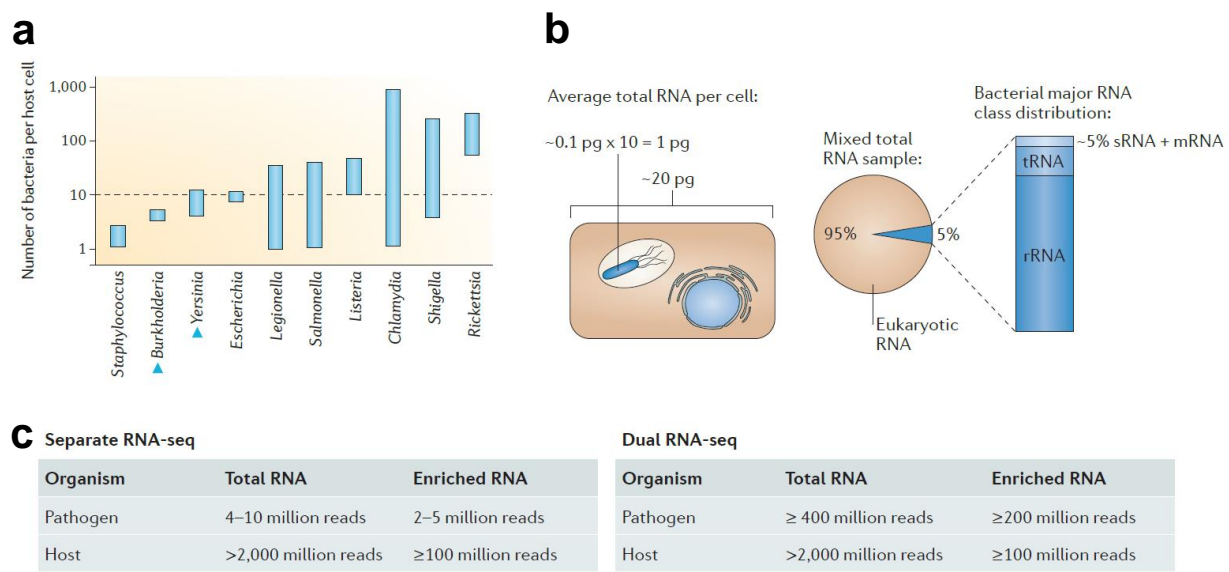


Figure 1.5|Estimation of the minimal sequencing depth requirements for Dual RNA-seq. a, Common infection models of bacterial pathogens were screened for reported numbers of individual pathogens per host cell. In these, a given host cell typically interacted with approximately ten bacterial cells (dashed line). **b**, Total cellular RNA content of a prototypical HeLa cell exceeds that of a prototypical bacterial pathogen by approximately a factor of 200. Based on the assumption that on average each host cell would be populated by ten bacterial cells (see panel a), this number decreases to factor 20 (= 5%). Informative RNA classes (mRNAs and regulatory non-coding RNAs) account for ~5% of both a mammalian and a bacterial transcriptome. Thus, only 5% of 5% (= 0.25%) of the total RNA isolated from an infected cell are expected to represent bacterial mRNAs and sRNAs and 5% of 95% (= 4.75%) to account for host mRNAs, lncRNAs and miRNAs. **c**, Based on these fractions and current sequencing depth recommendations for single-organism RNA-seq (see main text), the required depth for Dual RNA-seq was extrapolated. Figure taken from Westermann et al., 2012.

To assess the feasibility of Dual RNA-seq in a “standard” infection model of mammalian cells and bacterial pathogens, the average numbers of bacteria per infected host cell were compiled from the literature (Fig. 1.5a). Based on the above mentioned total RNA content of an average eukaryotic or prokaryotic cell, the relative proportion of bacterial transcripts in a mixed host-pathogen RNA sample was estimated to be around 5% (Fig. 1.5b). Of these, again only ~5% are expected to reflect informative transcript classes such as mRNAs or sRNAs. Aiming at the suggested minimum of 1 M non-rRNA reads for bacterial RNA-seq (Haas et al., 2013), a total of 400 M reads would be needed for a Dual RNA-seq experiment based on total RNA to capture pathogen gene expression (Fig. 1.5c). Likewise, ~95% of the sampled RNA would be eukaryotic and rRNA would again dominate. Calculating with 5% non-rRNA transcripts and aiming at the recommended sequencing depth for mammalian RNA-seq of ~100 M (Toung et al., 2011), one ends up with 2,000 M reads being required for a Dual RNA-seq experiment to capture host gene expression (Fig. 1.5c). This indicates that counter-intuitively the host would be the limiting factor in the given setting. Note, however, that the predicted proportions of non-rRNA molecules

in the sample might not accurately reflect the actual proportion of reads derived from these classes in the final cDNA library. It is thus concluded that – even though the precise sequencing depth would have to be defined empirically – Dual RNA-seq for an infection model based on mammalian cells challenged with bacterial pathogens appears feasible using current technology (e.g. by Illumina sequencing).

1.7 Aim of the study

The present study was sought to establish the Dual RNA-seq approach – i.e. the simultaneous analysis of host and pathogen gene expression during infection – for the model system of human cell lines and intracellular *S. Typhimurium*. To this end, an experimental protocol might be based on the infection of host cells with constitutively green fluorescent protein (GFP)-expressing *Salmonella* to be able to distinguish invaded host cells from non-infected ones. Infected cells should be fixed, invaded cells and non-infected bystanders be separated via fluorescence-activated cell sorting (FACS), total RNA extracted and cDNA libraries constructed. These would subsequently be analyzed by Illumina sequencing. From the resulting data, novel *Salmonella* virulence genes may be identified and their biological role in the adaptation of the pathogen to its intracellular lifestyle and corresponding consequences in host gene expression be unraveled. The overall goal was to develop a tool that would allow infection biologists to assess the global picture of an infection event and ultimately, to link given virulence phenotypes to the underlying molecular mechanisms.

2. Implementation of the Dual RNA-seq approach

This chapter describes a detailed protocol for Dual RNA-seq of two very different interaction partners, the bacterial pathogen *S. Typhimurium* and human epithelial or monocyte-like cells. Therefore, the present study may be considered a generic description of the application of Dual RNA-seq to virtually any organisms – irrespective of kingdom affiliation – which are amenable to also conventional RNA-seq. Note however, that the relative proportion of host to pathogen reads may differ for other model systems and the required sequencing depth itself largely depends on the respective scope of a study.

2.1 Choice of the infection model and optimization of infection conditions

In order to evaluate the general applicability of Dual RNA-seq for bacterial/mammalian host-pathogen interactions we established an *in vitro* infection model based on human cervical carcinoma HeLa-S3 cells and *S. Typhimurium* strain SL1344. This model represents a robust, widely accepted and long studied assay system (Kihlstrom, 1977). Infection of epithelial cells involves SPI-1-dependent invasion which is followed by a phase of SPI-2-dependent intracellular replication (Galan and Curtiss, 1989; Garcia-del Portillo et al., 1993). Furthermore, the HeLa-S3 cell clone has recently been suggested by the ENCODE consortium as a preferred model cell line for various experimental setups and specific culture conditions were proposed to ensure comparability between different studies (Birney et al., 2007; Dunham et al., 2012). Its selection will therefore enable a direct comparison of our infection data with a wealth of additional experimental conditions. To discriminate between invaded host cells and non-infected bystanders a constitutively GFP-expressing *Salmonella* strain (Papenfort et al., 2009) is used. In this strain the coding sequence of *gfp* is placed downstream of the constitutive p_{tet} promoter and integrated into the *put* locus within the *Salmonella* chromosome.

A scheme of the experimental pipeline from the infection assay to RNA-seq is given in Figure 2.1: HeLa-S3 cells are cultured according to the ENCODE guidelines and infected with *S. Typhimurium* following a previously published protocol from our group (Schulte et al., 2011). That is, cells which were seeded 2 d in advance are challenged with defined doses of *S. Typhimurium* that were grown to an OD₆₀₀ of 2. Subsequently, the cells are centrifuged briefly to mediate pathogen/host cell contact and synchronize the infection event. After incubation for 30 min at 37°C to allow invasion to occur, the remaining extracellular pathogens are killed by the addition of gentamicin. Invaded bacteria, however, are protected due to the host membrane

impermeability for the antibiotic. Samples are taken at defined time points after infection and fixed (see below). FACS-sorting is used to separate invaded and bystander cells, leading to the enrichment of *Salmonella*-infected cells and the resolution of the specific responses of extra- and intracellularly stimulated hosts. For a pilot experiment, RNA standards at a fixed concentration are added to the biological sample as an internal quality control ('spike-in RNA'; see page 40). A sonication-based shearing step, that fragments RNAs into near uniform-length sizes in the range of 200-400 bp, is included to avoid transcript length biases during the downstream PCR amplification. Ligating RNA adapters to the fragmented transcripts in a defined orientation ensures that the resulting cDNA libraries are strand-specific. The barcode sequences used for multiplexing are introduced during PCR. The final cDNA libraries are sequenced on Illumina's HiSeq 2000 platform (100 cycles; single-end mode) yielding ~100-150 M reads per lane.

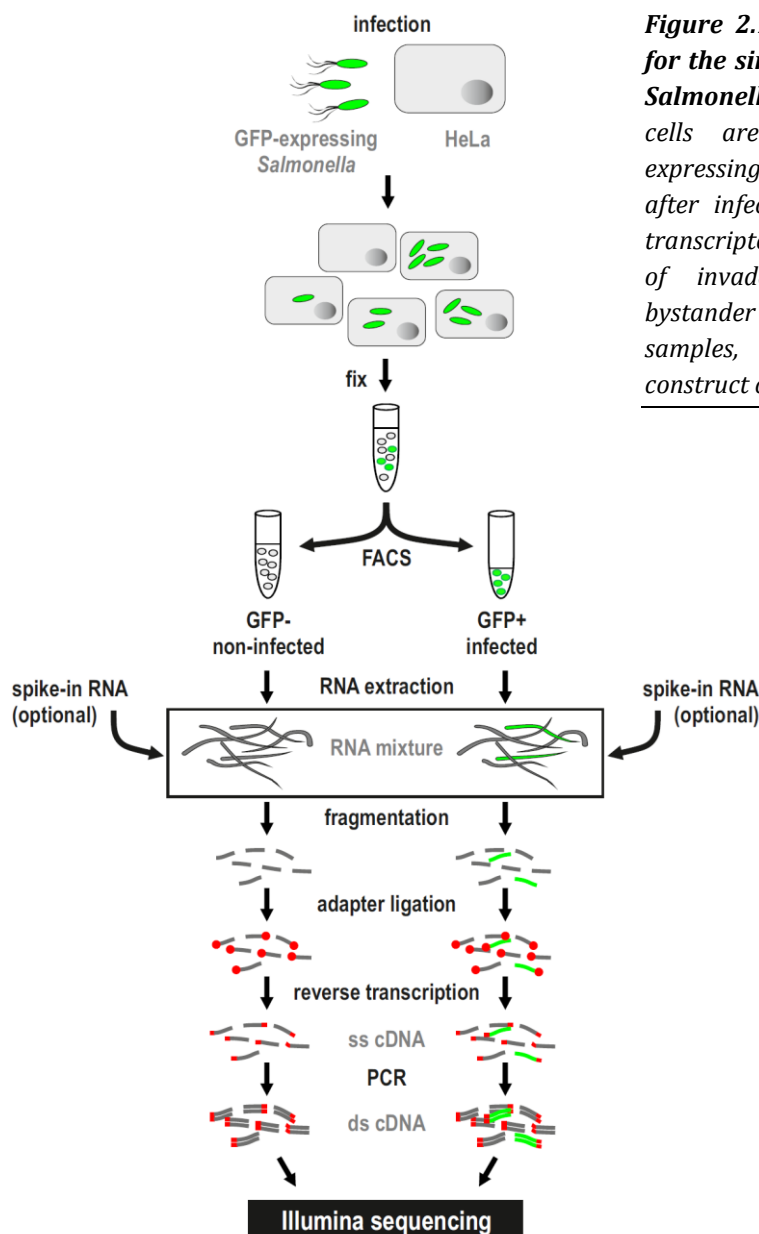


Figure 2.1|Experimental workflow established for the simultaneous transcriptome analysis of *Salmonella*-infected mammalian cells. HeLa-S3 cells are infected with constitutively GFP-expressing *S. Typhimurium*. At a given time point after infection, cells are fixed to preserve their transcriptome and sorted into the sub-populations of invaded (GFP-positive) and non-infected bystander cells (GFP-negative). From the resulting samples, total RNA is extracted and used to construct cDNA libraries.

Initially, HeLa-S3 cells were infected with a multiplicity of infection (m.o.i.) of 10, an infectious dose frequently used in the literature (e.g. Sittka et al., 2007). A thus infected HeLa culture was visualized by confocal laser scanning microscopy (CLSM) at 4 h post-infection (p.i.) (Fig. 2.2). This revealed that challenging HeLa-S3 with *S. Typhimurium* gives rise to a mixed population of infected (GFP-positive) and non-infected (GFP-negative) host cells, with the former usually containing more than just a single pathogen presumably reflecting multiple infection events. As predicted, gentamicin selectively killed extracellular pathogens (highlighted in the control by white arrows) whereas intracellular bacteria were unaffected.

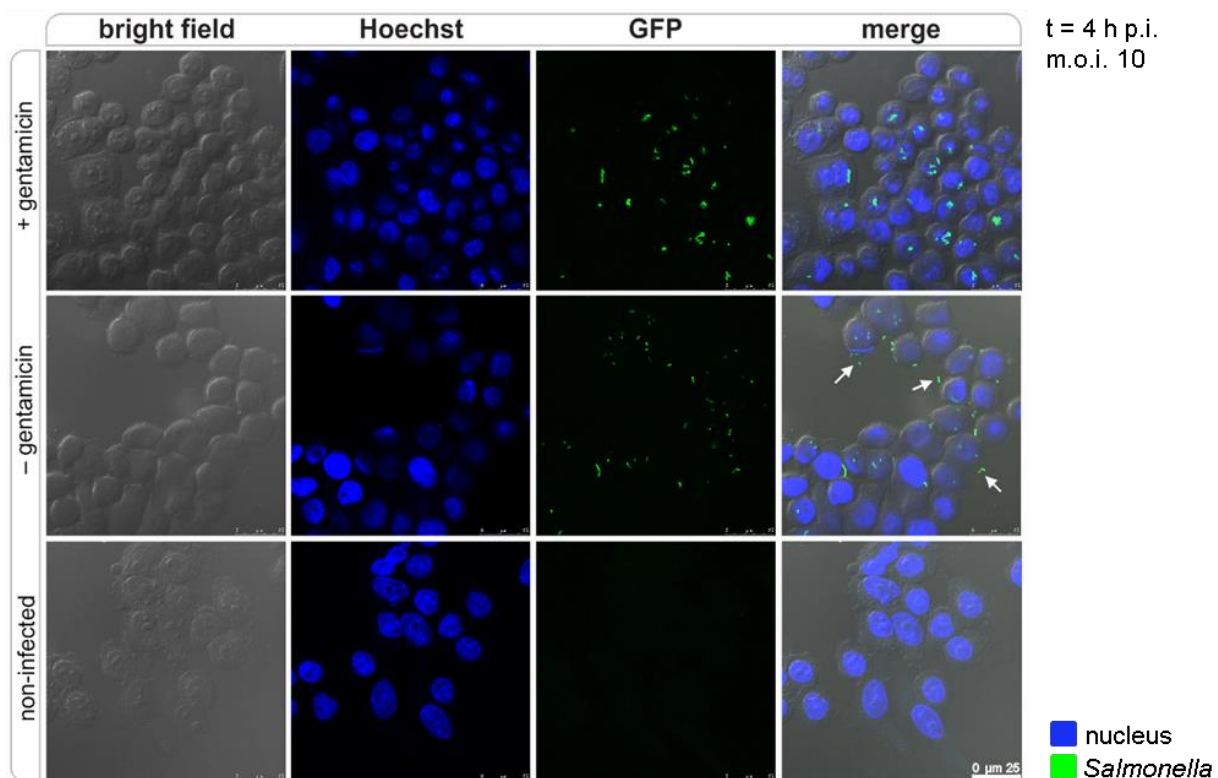


Figure 2.2|Fluorescence microscopy of HeLa-S3 infected with GFP-expressing *S. Typhimurium*. Cultures of HeLa-S3 cells were either infected with wild-type *Salmonella* (m.o.i. 10) or left uninfected. At 4 h p.i., chromosomal DNA was stained with the Hoechst dye and the cells were visualized by CLSM. The scale bar indicates 25 μ m.

To obtain detailed information on the respective fates of host and pathogen after infection, HeLa-S3 infectivity and cell death were monitored over a range of increasing bacterial doses (Fig. 2.3) and intracellular replication of invaded pathogens was traced over time (Fig. 2.4). On the host side, the infection rate largely depended on the m.o.i. as revealed by GFP-based flow cytometric analysis (Fig. 2.3a). When comparing the rate of infected HeLa-S3 cells for discrete m.o.i. values between 4 and 24 h p.i., a reduction in the relative amount of infected cells over

time was observed; in particular if high bacterial doses had been applied (m.o.i. ≥ 10). This discrepancy can be explained by an increase in cellular stress resulting in apoptosis (Fig. 2.3b) and necrosis (Fig. 2.3c). For the purpose of analyzing infection-associated gene expression patterns, unspecific processes (such as necrosis) should be avoided. As a consequence, a m.o.i. of 5 was selected for further assays, thereby mostly preventing the induction of cell death upon infection.

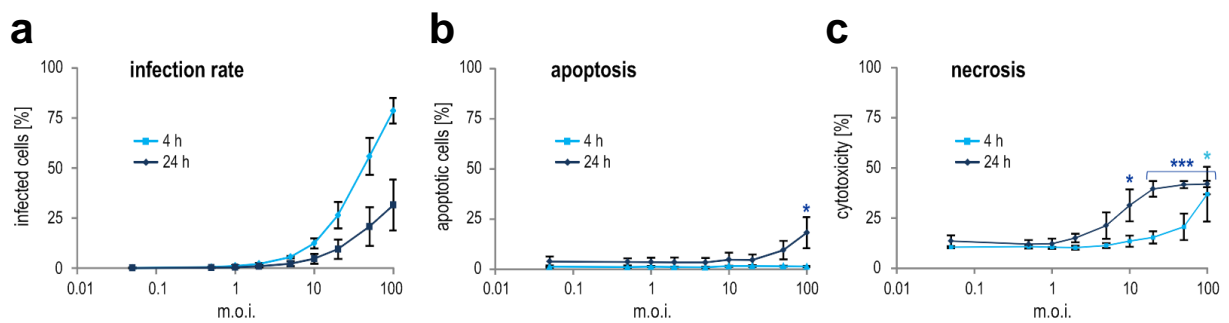
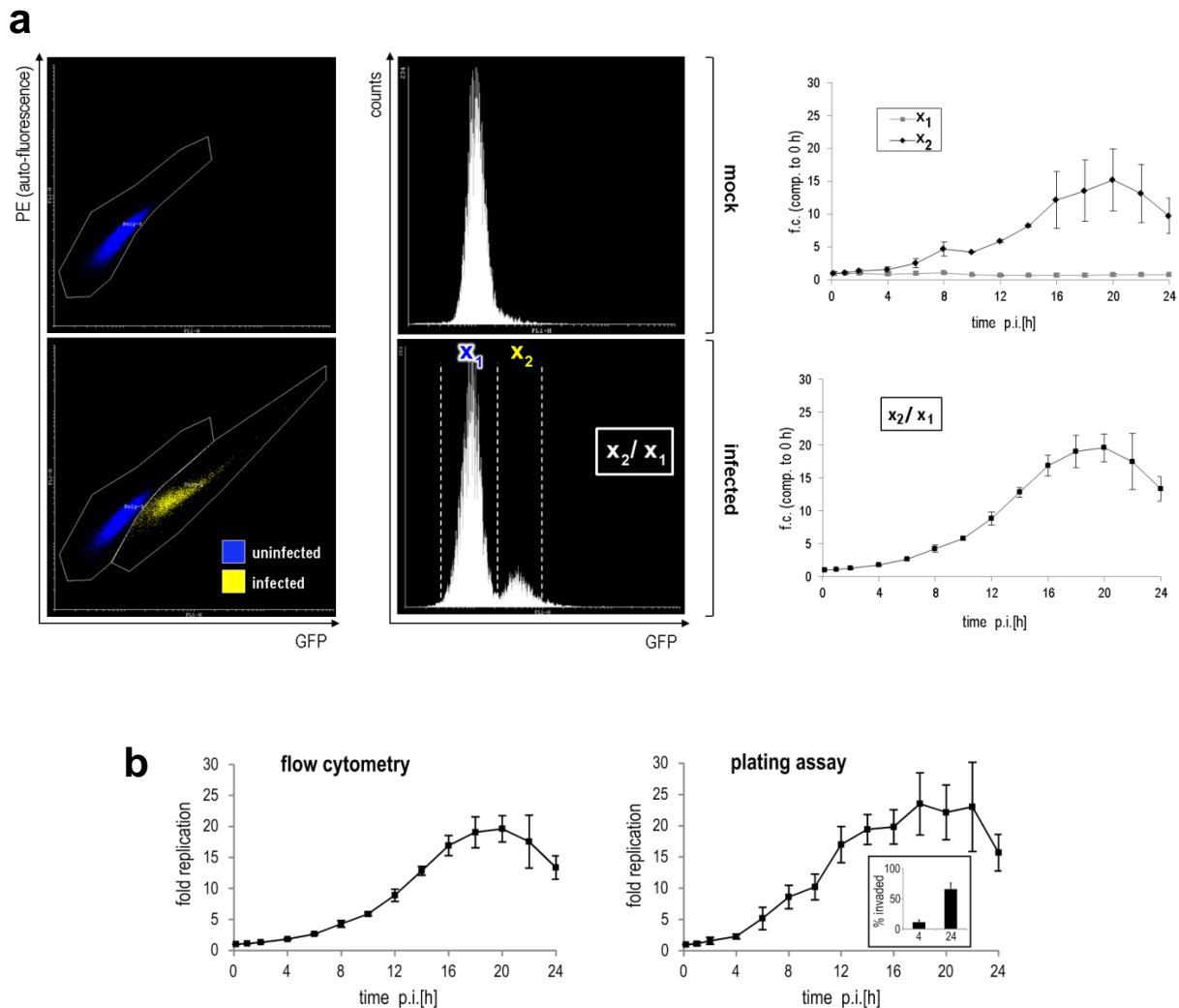


Figure 2.3/Measuring bacterial dose effects on host cell infectivity, apoptosis and cytotoxicity. HeLa-S3 cells were infected with increasing doses of GFP-expressing *S. Typhimurium* (m.o.i. 0.05-100) for 4 h or 24 h. **a**, The infectivity rate was determined by flow cytometry based on a positive GFP signal for infected cells. **b**, The rate of apoptotic cells was measured after Annexin V/propidium iodide staining by flow cytometry. **c**, Host cell cytotoxicity was determined by measuring the release of the cytosolic enzyme lactate dehydrogenase (LDH) into the culture medium as a proxy for cell death. The maximum amount of cytotoxicity was defined by treating the cell culture with lysis buffer prior to the analysis. *: $p < 0.05$; ***: $p < 0.001$ (as compared to non-infected cells; one-tailed Student's *t*-test). The data reflect the mean \pm SD from biological triplicates.

To trace intracellular replication, we developed a novel protocol for the rapid quantification of bacterial growth inside mammalian host cells (Fig. 2.4a): Briefly, infected host cells were analyzed by flow cytometry as described above, except that GFP intensity (rather than the relative proportion of GFP+ cells) was quantified in the invaded sub-population over time and compared to that immediately after infection ($t = 0$ h). In particular, the geometric mean in the histogram (GFP channel) for the GFP+ fraction was normalized against that of the non-infected fraction. The increase in normalized GFP intensity over time was used as a proxy for bacterial proliferation (see Methods section for further details). As shown in Figure 2.4b, the obtained kinetics matched well to those obtained from conventional plating assays. If infection was carried out at a m.o.i. of 5, intracellular replication of invaded bacteria was initially observed from ~ 4 h p.i. and reached its plateau around 20 h p.i. The subsequent decrease might be due to active killing by the host or – at least to some extent – reflect residual traces of host cell death even at this low bacterial dosage as indicated in the cytotoxicity measurements (Fig. 2.3c).

Finally, based on the determined infection rate (Fig. 2.3a) with a m.o.i. of 5 at 4 h (5.48%) or 24 h p.i. (2.17%), respectively, and the amount of colony-forming units (c.f.u.) recovered from lysed host cells (Fig. 2.4b, inset), the average number of intracellular bacteria per HeLa cell at these two time points was estimated: 4 h p.i., i.e. at the time when intracellular replication is about to start, there were on average ten bacteria within each invaded cell. This number increases to ~75 bacteria per infected cell at 24 h p.i.



Together, the *in vitro* infection model of ENCODE-compatible HeLa-S3 cells and constitutively GFP-expressing *S. Typhimurium* was characterized into great detail and demonstrated to generate very reproducible results. Moreover, as a consequence of the fluorescent *Salmonella* strain used, gene expression changes in the two sub-populations of invaded and bystander host cells can be separately monitored. Finally, due to the selection of a relatively mild infectious dose and the resulting minimization of host cell death, the model reflects physiological conditions. It should therefore enable the identification of infection-associated gene expression programs in both organisms and consequently, shed light on the intimate interplay between host and pathogen on the transcriptomic level.

2.2 Transcriptome fixation

Infection under the conditions used above results in a rather small proportion of invaded (GFP-positive) cells. GFP-based FACS can be applied to enrich this sub-population. As sorting is a time-consuming process, however, cells should be fixed beforehand to freeze the transcriptomes at the time when the samples are taken.

A suitable preservative for the given objective would meet two criteria. First, fixation should not affect RNA integrity/quality as a prerequisite for downstream analysis by RNA-seq and second, it should not interfere with GFP fluorescence as this parameter is the basis for FACS-based discrimination of host populations. Here, seven commonly used fixation methods were tested with regard to these two parameters (Fig. 2.5): Fixation with different concentrations of ethanol (EtOH) (Gillespie et al., 2002), the addition of ‘stop solution’ (i.e. 95% EtOH/5% phenol) (Eriksson et al., 2003), using the $(\text{NH}_4)_2\text{SO}_4$ -based reagent *RNAlater* (Qiagen), or different concentrations of paraformaldehyde (PFA) with or without 2% sucrose (W.-D. Hardt, personal communication). RNA quality of samples derived from these fixed cells was judged by capillary electrophoresis and compared to an unfixed control (Fig. 2.5a). Interestingly, the majority of fixation methods resulted in a characteristic fragmentation pattern of total RNA in accordance with previous findings (Cox et al., 2006). For instance, EtOH-mediated fixation led to a concentration-dependent increase in the fragmentation of the major eukaryotic ribosomal transcripts. In the given experimental setup, this would not necessarily be disadvantageous as the RNA samples are subsequently further fragmented prior to library generation. However, in addition to the concentration-dependency a time-dependent effect on fragmentation was observed (Fig. 2.5b). Due to this high degree of variability, we refrained from EtOH fixation. PFA treatment on the other hand, induces unspecific cross-linking of cellular RNA, DNA and protein. As a consequence, without reversal of the cross-link no RNA could be extracted (Fig. 2.5a). In contrast, when cross-linking was reverted by heating the samples for 15 min to 70°C as

proposed previously (Hamatani et al., 2006; Kuramochi et al., 2006), normal amounts of RNA could be isolated, however, the quality of such samples was severely impaired (Fig. 2.5a). Treatment with stop solution – which was developed for prokaryotic transcriptomes (Udekwu et al., 2005) – also negatively affected RNA quality. The only preservative that did not affect RNA integrity was RNAlater.

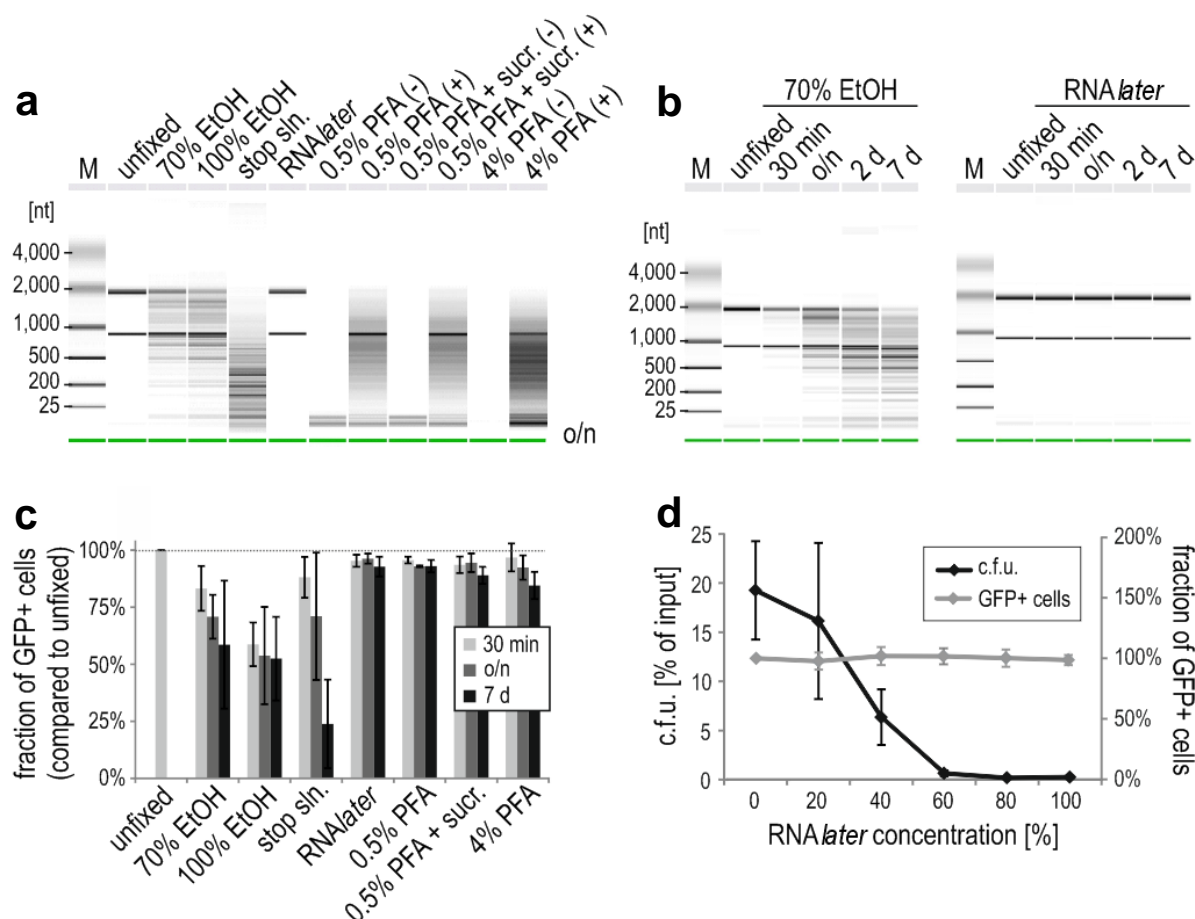


Figure 2.5|Evaluation of different transcriptome fixation techniques. *HeLa-S3* cells were infected with *S. Typhimurium* (m.o.i. 5) and cell samples were taken 4 h p.i. and analyzed. **a**, The influence of various preservatives on RNA integrity was evaluated using capillary electrophoresis for RNA samples extracted from differentially fixed samples. EtOH: ethanol; PFA: paraformaldehyde; suc.: sucrose (2%). For PFA-fixed samples, the induced cross-link either was (+) or was not (-) reversed by heating the samples for 15 min at 70°C before RNA extraction. In all cases fixation was performed overnight. Note that only eukaryotic rRNA bands are visible as the bacterial ones were too faint within the host background. **b**, Influence of the fixation period on RNA integrity. Cell samples were fixed in either EtOH or RNAlater for the indicated periods and RNA was analyzed as in panel a. **c**, The influence of fixation on GFP fluorescence was evaluated by determining the relative fraction of GFP-positive cells at different time points upon fixation (compared to unfixed samples). **d**, Increasing concentrations of RNAlater (diluted in PBS) reduce the amount of viable bacteria (black curve). The relative fraction of detectable GFP-positive cells is concentration-independent (grey). The data represent the mean from biological triplicates and error bars indicate +/- SD.

Furthermore, in contrast to a recently published report (Zaitoun et al., 2010), *RNAlater* did not seem to interfere with GFP fluorescence (Fig. 2.5c). In our hands, even when storing the cell samples for one week in *RNAlater*, the relative amount of detectable GFP-positive cells was unchanged. This does not hold for other fixation reagents which led to a decrease in the fraction of GFP-positive cells over time, such as EtOH (time- and concentration-dependent quenching in accordance to previous findings (Becker et al., 2012)) or stop solution (time-dependent quenching; likely due to the phenol contained in the mixture).

To assess whether different fixation protocols influence the relative representation of major (pro- and eukaryotic) transcript classes, RNA samples analyzed in Figure 2.5a were converted into cDNA and sequenced. Indeed, the relative proportion of individual transcript classes was dependent on the fixation method used (Fig. 2.6), especially for the bacterial transcriptome (upper panel). EtOH treatment for instance, led to a marked reduction in the representation of bacterial ribosomal transcripts, while tRNAs accumulated upon fixation. Also mRNAs were slightly depleted upon both EtOH and PFA fixation. Surprisingly, despite the striking effects of most of the fixatives on human rRNA (in Fig. 2.5a,b bacterial ribosomal bands are not visible due to the huge host background), the resulting human RNA-seq profiles were much more uniform than their bacterial counterparts. In summary, *RNAlater* again performed best as both the human and the bacterial RNA-seq profiles were virtually identical to those from the unfixed control.

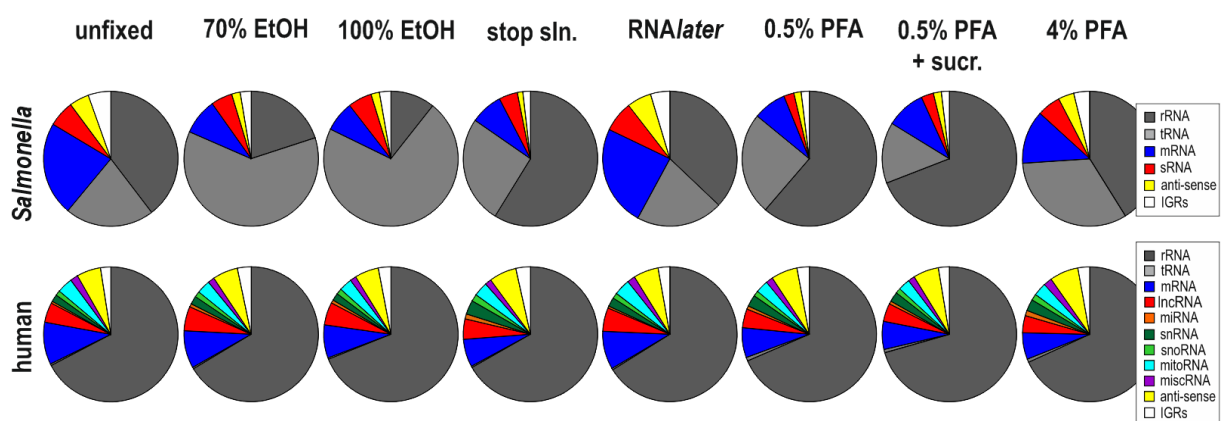


Figure 2.6/Global profile of *RNAlater*-fixed transcriptomes highly resembles those from unfixed cells. RNA samples derived from infected and differentially fixed HeLa-S3 cells (analyzed in Fig. 2.5a) were sequenced on Illumina's HiSeq 2000 platform to ~10 M reads/sample.

It was further determined if *RNAlater* treatment of infected host cells would also fix intracellular bacteria. As a proxy for effective fixation of invaded pathogens, the concentration of *RNAlater* that sterilized the sample was determined. Apparently a concentration of $\geq 80\%$ was sufficient to

abolish growth of bacteria recovered upon host cell lysis (black graph in Fig. 2.5d). This indicates that when using *RNAlater* in this or higher concentration, the reagent perfuses the host cell and fixes the intracellular bacteria. In parallel, the potential influence of the *RNAlater* concentration on the detectable fraction of GFP-positive cells was evaluated (grey line in Fig. 2.5d), but virtually no effect on this parameter was found. Together, for Dual RNA-seq analyses – at least in the context of the described infection model – *RNAlater* appears to be the best suited preservative as it was found to neither affect GFP fluorescence (Fig. 2.5c) nor RNA quality (Fig. 2.5a,b) and is compatible with RNA-seq (Fig. 2.6). Note however, that *RNAlater* fixation is not necessarily the optimal choice for any infection model. With other cell types (especially certain primary cells or tissue samples), *RNAlater* treatment might be too harsh and eventually cause cell lysis. Thus, it is recommended to evaluate further fixation protocols in the future, especially in the context of systems based on primary cells and *in vivo* models. Note further that *RNAlater*-fixed cells tend to clump. Thus, filtering the samples prior to sorting is essential in order to prevent from clogging the cell sorter.

2.3 FACS-based enrichment of infected cells

The fixed samples consisting of invaded (GFP-positive) and non-infected bystander cells (GFP-negative) were subjected to FACS analysis. A typical FACS plot is depicted in Figure 2.7a (panel ‘pre-sort’). In order to prevent cross-contamination, a conservative gating strategy was selected (i.e. leaving a gap between the two gates). To evaluate the purity of the thus sorted fractions, the resulting samples were re-analyzed using the same settings as prior to the sort (Fig. 2.7a, panel ‘re-analysis’ and 2.7b). This revealed that the GFP-positive fraction also contained a small fraction of GFP-negative cells (~10-15%). Contaminations might be due to cell duplets consisting of GFP-positive and -negative cells that – as a consequence of the GFP signal detected – are sorted in the GFP-positive collection tube. Duplet discrimination, however, in this case is not useful as cell shape, size and granularity are heterogeneous and largely depend on the bacterial load per host cell and thereby singlets may not be easily distinguishable from duplets. Alternatively, GFP-negative cells in the GFP-positive fraction might have initially (i.e. pre-sort) contained intracellular bacteria but due to the mechanical forces applied during sorting, the plasma membrane might have been ruptured and the bacterial load been released. This is likely to be the case especially for the 24 h time point, when infected cells are densely packed with about 75 individual bacteria (see above) and are thus fragile. Consequently, at least some of such cells might not even be considered true contaminations at all.

The GFP-negative sub-population on the other hand appeared highly pure (Fig. 2.7a, panel ‘re-analysis’). To test whether this fraction indeed represents non-infected bystander cells, the ratio

of bacterial *rfaH* mRNA vs. human β -actin mRNA was quantified in sorted sub-populations (24 h p.i.) by qRT-PCR analysis (Fig. 2.7c). Indeed *rfaH* mRNA was hardly detectable in GFP-negative cells. As inferred from these data, GFP-positive samples are enriched for bacterial RNA by approximately 500 fold as compared to GFP-negative cells. Taken together, as supported by two independent methods, GFP-based cell sorting results in highly enriched sub-fractions of invaded and non-infected host cells, originating from the same starting culture.

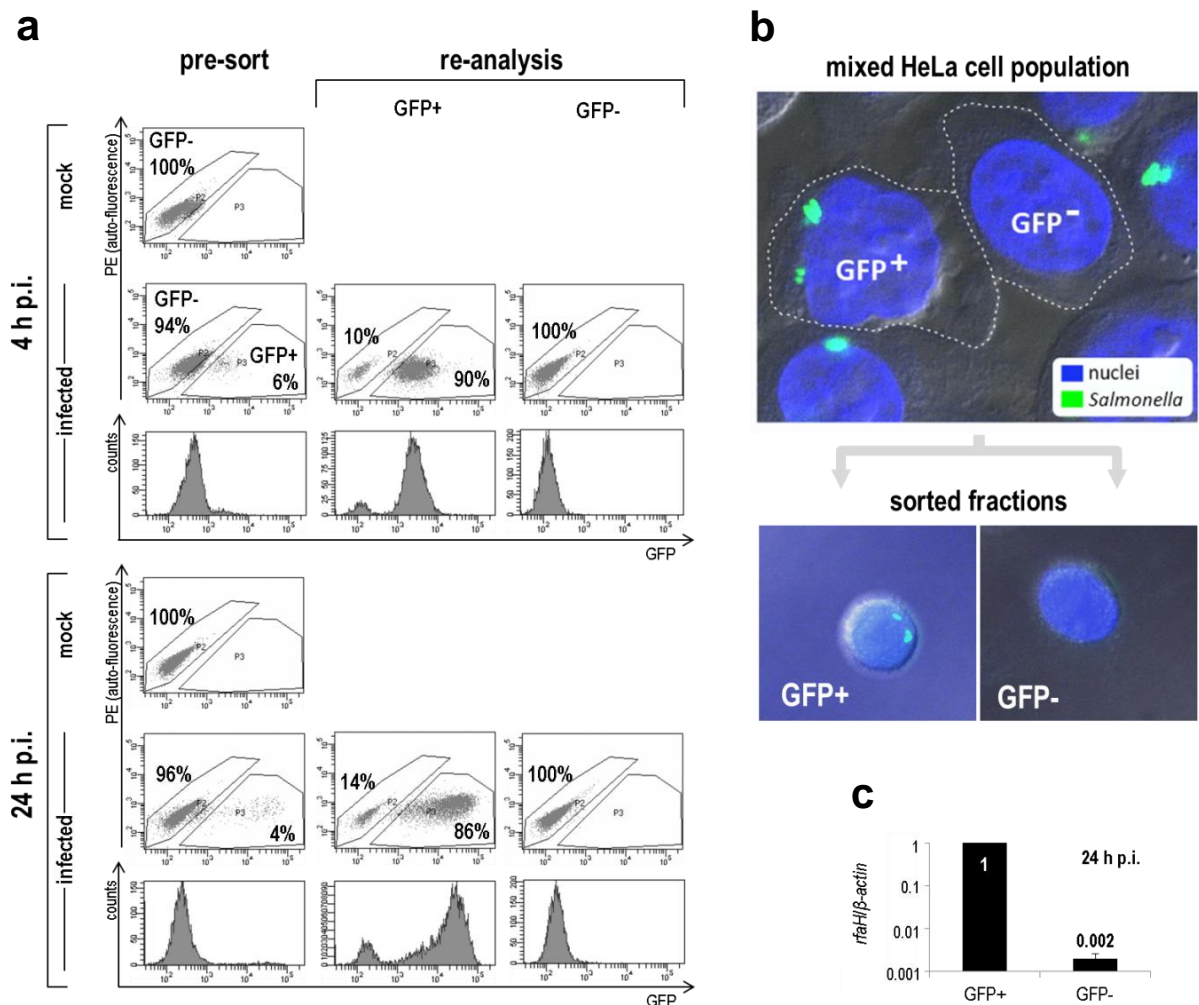


Figure 2.7|GFP-based separation of infected from non-infected HeLa-S3 cells by cell sorting results in a high enrichment of the sub-populations. a, Representative FACS plots of *Salmonella*-infected HeLa-S3 at 4 h or 24 h p.i. (m.o.i. of 5). The percentage values indicate the relative proportion of GFP+ and GFP- cells prior to and after sorting. **b**, Representative fluorescence microscopy images of HeLa-S3 cultures infected for 4 h prior to or after FACS enrichment. **c**, The purity of sorted cell populations was further assessed by qRT-PCR on extracted total RNA samples from cells that had been infected for 24 h and were FACS-enriched. *Salmonella rfaH* mRNA serves as a proxy for the bacterial and β -actin mRNA for the human transcriptome. The mean \pm SD from biological triplicates is shown.

The relative proportion of the bacterial transcriptome within the host background at distinct time points after infection was determined either prior to or after FACS-based enrichment of invaded cells (Fig. 2.8). For this purpose *rfaH* and β -actin mRNAs were again used as a proxy for the bacterial or human transcriptome, respectively. Because it was unknown whether these two transcripts are being expressed to a similar level in the context of their respective transcriptome, a dilution series of separately isolated *S. Typhimurium* and human total RNA was prepared (Fig. 2.8a). From the resulting trend-line equation the relative proportion of bacterial RNA within an infectious sample at different time points p.i. could be deduced (Fig. 2.8b, blue bars). Based on this estimation the relative proportion of the *Salmonella* transcriptome increases by ~ 7 fold from 4 h to 24 h p.i. (both for sorted and unsorted samples), reflecting intracellular replication kinetics (Fig. 2.4). In sorted fractions, bacterial RNA constitutes an estimated proportion of between $\sim 0.3\%$ (4 h p.i.) and $\sim 2\%$ (24 h p.i.) of the total RNA sample.

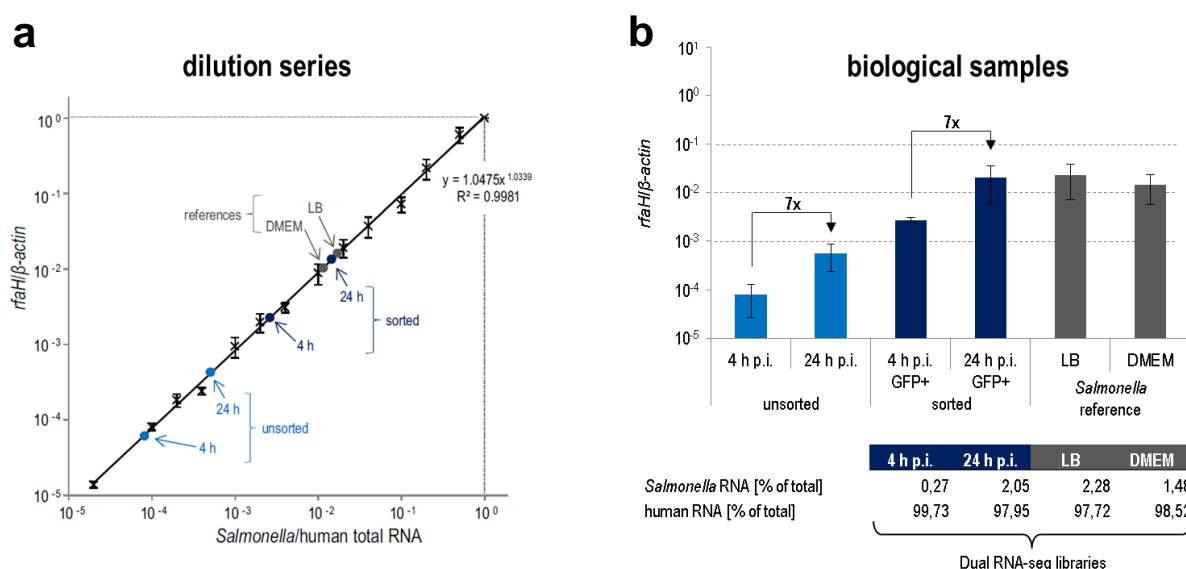


Figure 2.8/Estimation of the fraction of bacterial RNA in the mixed host-pathogen transcriptome samples. **a**, A serial dilution of separately isolated *Salmonella* to HeLa-S3 total RNA was prepared and bacterial *rfaH* (relative to human β -actin) mRNA was quantified by qRT-PCR. **b**, The resulting trend-line equation from panel a was used to deduce the percentage of the *Salmonella* transcriptome within mixed total RNA samples from infected HeLa-S3 cells at different time points and with or without FACS-based enrichment of the invaded cells. Depicted is the mean \pm SD from three biological replicates.

2.4 Computational challenges

cDNA libraries were supposed to be sequenced on the Illumina HiSeq 2000 platform with ~ 100 cycles in the single-end mode. A bioinformatics pipeline, based on the recently published READemption tool (Förstner et al., 2014), was compiled to meet the requirements of Dual RNA-seq (Fig. 2.9). That is, raw reads were first quality-filtered and reads < 20 nt were removed. The

remaining reads were next mapped simultaneously to the *Salmonella* and human reference genomes or the spike-in RNA sequence. Reads that could not be mapped to either reference sequence with an accuracy $\geq 90\%$ were considered ‘not mappable’ and discarded. Likewise, reads that could be mapped to more than one reference genome were defined as ‘cross-mappings’ and removed. The remaining reads mapped to only one reference sequence. Note however, that also these reads might map to multiple loci (but within the same genome). If a given read was mapped to n positions in the same genome, each of the annotated n genes to which a read was aligned was attributed with $1/n$ reads. The aligned reads can next be used to (1) calculate coverage plots that are visualized using the Integrated Genome Browser (IGB) (Nicol et al., 2009), (2) identify alternative splice patterns as well as circular transcripts (Doose et al., 2013), and (3) to call differential gene expression. In case of the latter, gene-wise read quantifications are calculated and used as input data for the DEseq tool (Anders and Huber, 2010) to reveal gene expression changes between two conditions. Only transcripts that had been detected above a certain Reads per kilobase transcript per million mapped reads (RPKM) cut-off (see page 43) were included in the expression analyses. If applicable, differentially expressed genes can be put into a functional context via gene ontology (GO)-term analyses by analyzing KEGG (Kyoto Encyclopedia of Genes and Genomes) signaling pathways (Kanehisa et al., 2014) or using the Signal Pathway Impact Analysis (SPIA) tool (Tarca et al., 2009).

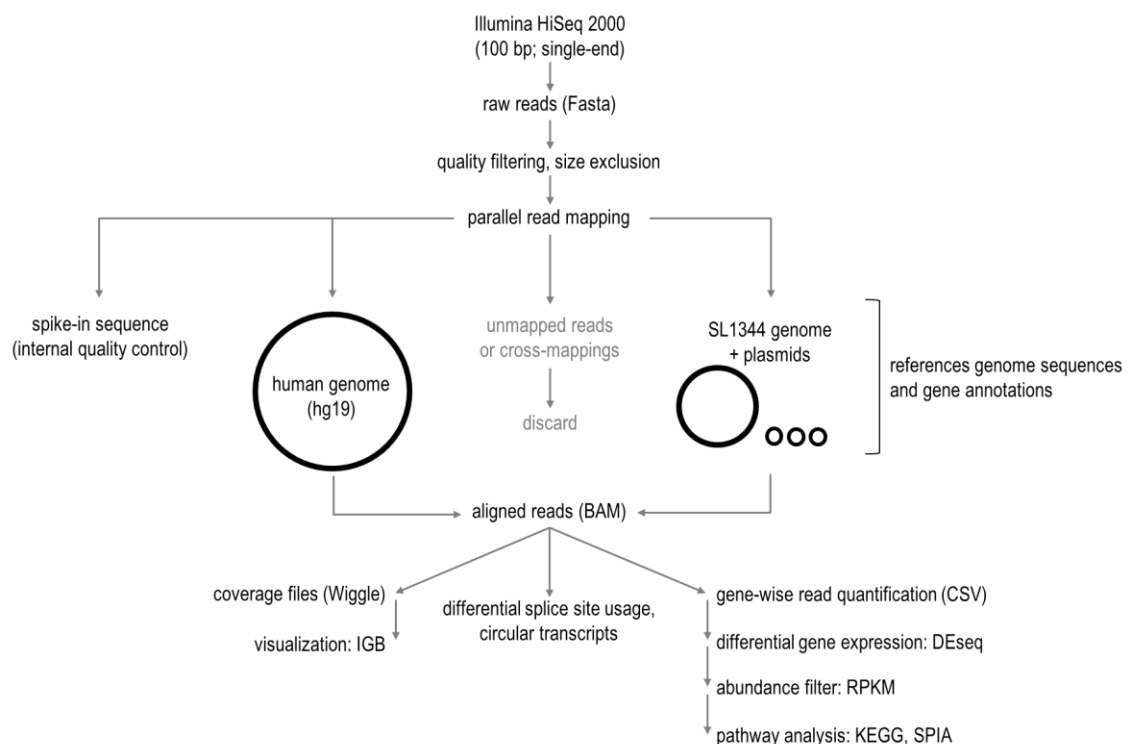


Figure 2.9|Computational work-flow for Dual RNA-seq. Most steps in the mapping pipeline are based on the READemption tool (Förstner et al., 2014). File types are given in brackets. Note that spike-in RNA standards were included only for pilot Dual RNA-seq.

2.5 Defining the minimal material requirements for a Dual RNA-seq experiment

The limiting steps in the Dual RNA-seq pipeline (Fig. 2.1) are prior to or during cDNA library generation. Due to the mild infectious dose and the resulting low infection rate (Fig. 2.3), collecting sufficient infected cells by cell sorting can be time-consuming. In order to be able to handle several samples in parallel the protocol had to be stream-lined. Therefore the minimal RNA amount required for Dual RNA-seq was empirically determined. GFP-positive cells were collected by FACS in decreasing amounts from 5×10^5 to 1×10^3 , total RNA was isolated and subjected to the standard Dual RNA-seq protocol (Fig. 2.1). As expected, a positive linear relationship between the input number of cells and the RNA output was observed (Fig. 2.10a, blue line). In contrast, the relative proportion of mappable reads followed a switch-like curve (red graph in Fig. 2.10a): Whereas for an input cell number of 10^4 ~75% of the raw reads could be mapped, this fraction sharply declined with lower input amounts. Starting from only 10^3 cells resulted in >70% of non-mappable reads. Irrespective of the initial RNA amount, all cDNA libraries were adjusted to the same concentration prior to sequencing. This means that samples derived from low input RNAs were more extensively amplified by increasing the number of PCR cycles. As a result, the cDNA libraries for these samples were biased towards smaller fragments (Fig. 2.10b). The smallest fragments ran at a height of ~100 nt. Note however, that these fragments already contained the poly(A) stretch that had been ligated prior to reverse transcription.

The quality of reads derived from low input samples decreased markedly at their 3' end (Fig. 2.10c). This might be an indication for homo-polymers, which are considered to hamper the Illumina sequencing reaction. In fact, the relative adenine content in the raw sequencing data (i.e. prior to poly(A) trimming) describes a similar switch-like curve as – although it anti-correlates with – the proportion of mapped reads (grey line in Fig. 2.10a): For samples derived from $\geq 10^4$ cells the relative fraction of adenines is around the expected 25% but it rapidly increases to up to 50% in less input samples. Hence, it seems as if for samples with <100 ng of input RNA the increasing numbers of PCR cycles would bias towards the smallest fragments of ~100 nt of which a considerable fraction represents 3' oligo(A) stretches. Despite poly(A) stretches being clipped computationally prior to the mapping process, many of the remaining sequences are presumed too short (i.e. <20 nt) to pass the length filter and thus, discarded (Fig. 2.9). Based on these findings, a minimal RNA amount of ~100 ng (corresponding to $\sim 1 \times 10^4$ infected HeLa-S3 cells) would be needed for Dual RNA-seq. These requirements should allow for the parallel processing of multiple samples.

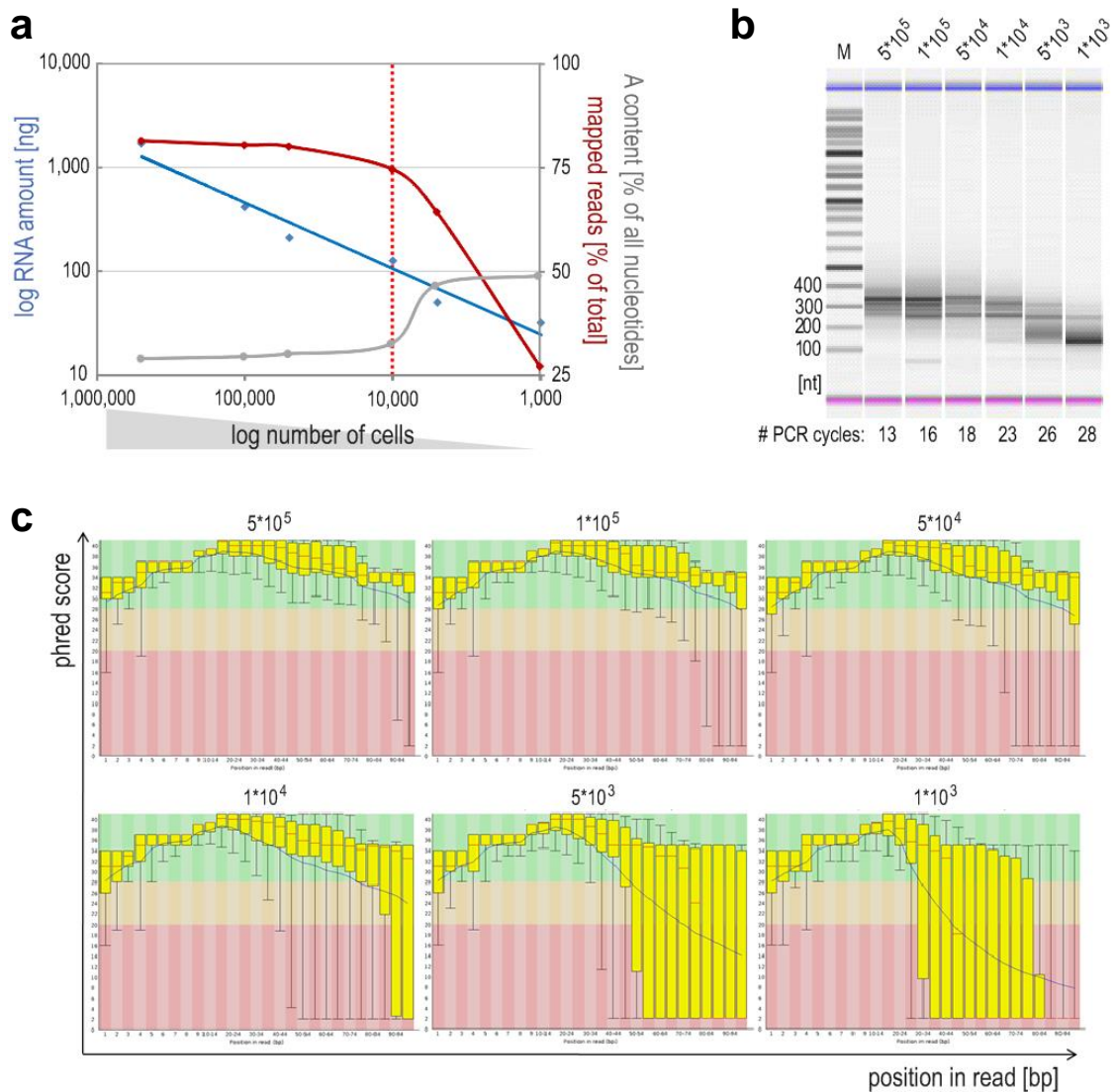


Figure 2.10/Minimal material requirements for Dual RNA-seq. *Salmonella*-infected HeLa-S3 cells were harvested at 4 h p.i. in decreasing amounts from 5×10^5 to 1×10^3 . Total RNA was extracted and the amount measured (blue graph in panel **a**; logarithmic scale). RNA was reverse transcribed and the resulting cDNA samples were PCR-amplified to the same amount (**b**). The resulting cDNA libraries were analyzed by capillary electrophoresis and pooled in an equimolar ratio prior to sequencing on Illumina's HiSeq 2000 (~20-30 M reads/library). The proportion of reads that could be mapped to the human or *Salmonella* reference genome as well as the proportion of adenine (A) nucleotides were determined for each sample (panel **a**; red or grey graph, respectively; linear scale). The red dotted line denotes 1×10^4 cells. **c**, Sequencing quality depends on the amount of input cells. For each sample, phred scores (as a measure of sequencing quality) were plotted on the y-axis and the relative position of the respective base within the 100 bp read is plotted on the x-axis. The decrease in phred scores at the 3' end anti-correlates with the proportion of A nucleotides (grey line in panel **a**) suggesting homo-polymers might hamper Illumina sequencing.

2.6 A proof-of-principle experiment

Three independent biological replicates from two defined time points after HeLa-S3 infection with *S. Typhimurium* (m.o.i. 5) were sequenced to moderate depth (i.e. ~20-35 M reads/sample). Each experiment was comprised of eight individual samples (Fig. 2.11; Tab. 2.1). These included two *in vitro* grown *Salmonella* reference controls from rich medium without ('LB') or upon medium shift from LB into the host cell medium and incubation therein for another 15 min at room temperature ('DMEM'). This was done as previous studies indicated that shifting bacterial cultures from LB to the host cell medium in which the infection assay will take place, affects global gene expression in the bacterium (e.g. Blair et al., 2013; Hautefort et al., 2008). For instance, due to the low iron content of mammalian cell culture media such as DMEM, numerous bacterial iron transporter genes are typically induced in tissue culture medium (Hautefort et al., 2008). Therefore, in order to be able to distinguish between gene expression changes associated with different culture conditions and true infection- and intracellular lifestyle-specific gene expression programs, both medium controls were included in the analysis.

A previous comparative study indicated that following common column-based extraction methods (e.g. via the RNeasy kit) short RNA species tend to become lost, whereas phenol-based reagents (e.g. TRIzol) bias against longer transcripts (Stead et al., 2012). TRIzol extraction has further been reported to deliver RNA in low yield for some bacterial species (Jahn et al., 2008). Thus, these methods seemed non-ideal for Dual RNA-seq experiments. Rather total cellular RNA was isolated using the *mirVana* RNA isolation kit – a procedure optimized to recover all types of RNA, including small and large RNA species.

Since the *mirVana* isolation protocol was optimized for mammalian cells it was necessary to prevent potential biases when extracting RNA from bacteria. Therefore, rather than isolating RNA from pure *Salmonella* samples, the ratio of bacteria to human cells was maintained by supplementing the *Salmonella* lysates with lysates from HeLa-S3 cells before RNA extraction (see Methods section). Early (4 h p.i.) and late infection time points (24 h p.i.) were taken, each of which was split into invaded cells and non-infected bystanders via FACS. As a reference control for the host side mock-infected HeLa cells were included for each time point and sorted for the GFP-negative fraction to ensure equal treatment between infected and naïve cells. Early time points were sequenced to an overall greater depth (~35 M reads/library) than the later ones (~25 M reads/library) to account for the expected lower fraction of *Salmonella* RNA in the early samples.

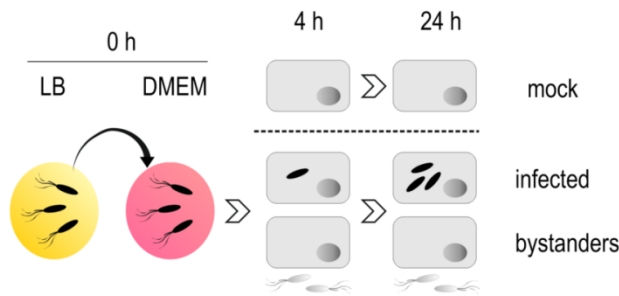


Figure 2.11|Experimental overview of pilot Dual RNA-seq. Two time points after *Salmonella* infection of HeLa-S3 cells were sampled. Infected host cells were separated from non-infected cells from the same culture by FACS. Mock-treated control cells at both time points sorted for the GFP-negative fraction served as a host reference. As pathogen controls, extracellular *Salmonella* were grown in LB to an OD_{600} of 2 and either were or were not shifted to host cell DMEM medium.

Table 2.1|cDNA libraries analyzed in the Dual RNA-seq pilot experiment.

#	library name	<i>Salmonella</i> strain	host cell type	time point	total reads	replicates
1	0 h WT LB	WT	HeLa-S3 (as carrier RNA)	0 h	~25 M	3
2	0 h WT DMEM	WT	HeLa-S3 (as carrier RNA)	0 h	~25 M	3
3	4 h mock GFP-	-	HeLa-S3	4 h	~25 M	3
4	4 h WT GFP+	WT	HeLa-S3	4 h	~25 M	3
5	4 h WT GFP-	WT	HeLa-S3	4 h	~25 M	3
6	24 h mock GFP-	-	HeLa-S3	24 h	~25 M	3
7	24 h WT GFP+	WT	HeLa-S3	24 h	~25 M	3
8	24 h WT GFP-	WT	HeLa-S3	24 h	~25 M	3
9	0 h WT LB	WT	THP-1 (as carrier RNA)	0 h	~25 M	1
10	0 h WT DMEM	WT	THP-1 (as carrier RNA)	0 h	~25 M	1
11	4 h mock GFP-	-	THP-1	4 h	~25 M	1
12	4 h WT GFP+	WT	THP-1	4 h	~25 M	1
13	4 h WT GFP-	WT	THP-1	4 h	~25 M	1
14	24 h mock GFP-	-	THP-1	24 h	~25 M	1
15	24 h WT GFP+	WT	THP-1	24 h	~25 M	1
16	24 h WT GFP-	WT	THP-1	24 h	~25 M	1

Even though technical reproducibility for RNA-seq was initially claimed to be high (e.g. Marioni et al., 2008; Mortazavi et al., 2008) more recent studies recommend the use of internal references that allow for inter-experimental comparison (e.g. McIntyre et al., 2011). For that reason biological samples were supplemented with synthetic spike-in RNAs (Jiang et al., 2011). In particular, two distinct spike-in RNAs with equal length but different GC-content (spike #1 and #2; see Methods section), were evaluated for their suitability in the Dual RNA-seq approach (Fig. 2.12). Both sequences were computationally tested to be absent from both the *Salmonella* and human genome. Both standards behaved linear within the tested range, i.e. an increased input concentration of spike-in RNA resulted in a higher read number mapped to the spike-in

sequence (Fig. 2.12). Importantly however, spike #1 (but not spike #2) sequestered only a minor fraction of the total sequencing reads meaning that the vast majority of reads was still mapped to the biologically relevant sequences (i.e. the two genomes). The relatively high occurrence of spike #2 might be due to the identity of its first nucleotide (a cytosine; see Methods) which has been considered a preferred substrate for the T4 RNA ligase that is used for library construction (Romaniuk and Uhlenbeck, 1983; Stark et al., 2006). In conclusion, spike-in #1 appears to be well suited for the given assay system and thus, was added to each of the RNA samples of the pilot experiment (see Fig. 2.1).

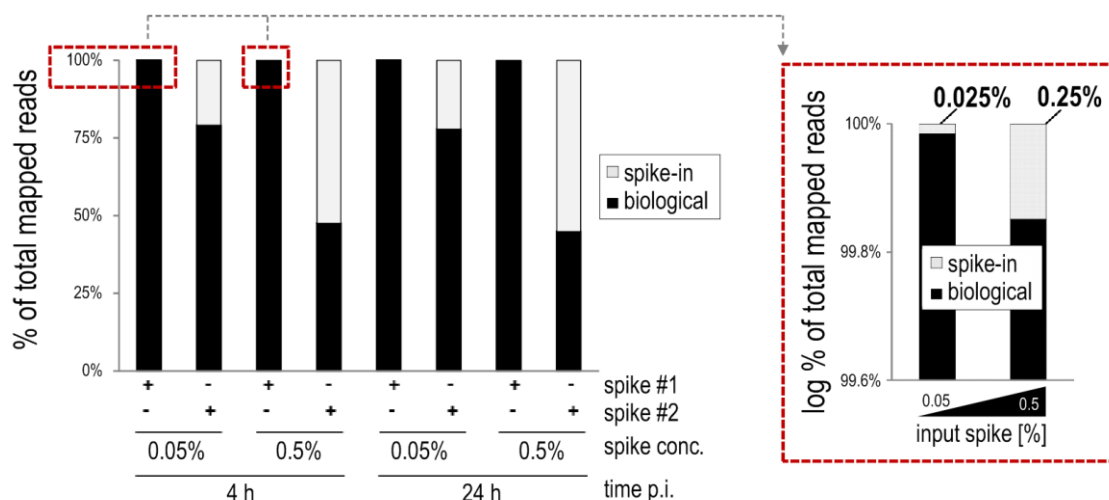


Figure 2.12/Spike-in RNA standards are used as an internal quality measure of Dual RNA-seq. Mapping statistics for total RNA samples that had been supplemented with different concentrations of 20 nt-long synthetic RNA standards (spike #1 and #2; see Methods section for details). Right panel: Magnification of the obtained read distribution for spike #1.

With a median length of ~96 nt (Fig. 2.13a), sequencing reads came close to the maximum of 100 nt. The quality-filtered sequencing reads from the two time points after infection were simultaneously mapped against the human (hg19 – hs37d5 annotation) and *Salmonella* (NC_016810.1) reference genomes, the three *Salmonella* virulence plasmids (NC_017718.1, NC_017719.1, NC_017720.1), as well as against the spike-in RNA sequence (spike-in #1). Only a minor fraction of reads could be mapped to both genome sequences and their fraction decreased with increasing mapping stringency (Fig. 2.13b). Based on these findings, an accuracy threshold of $\geq 90\%$ was applied and the ‘cross-mapped’ reads below that criterion were discarded and not included in the downstream analyses.

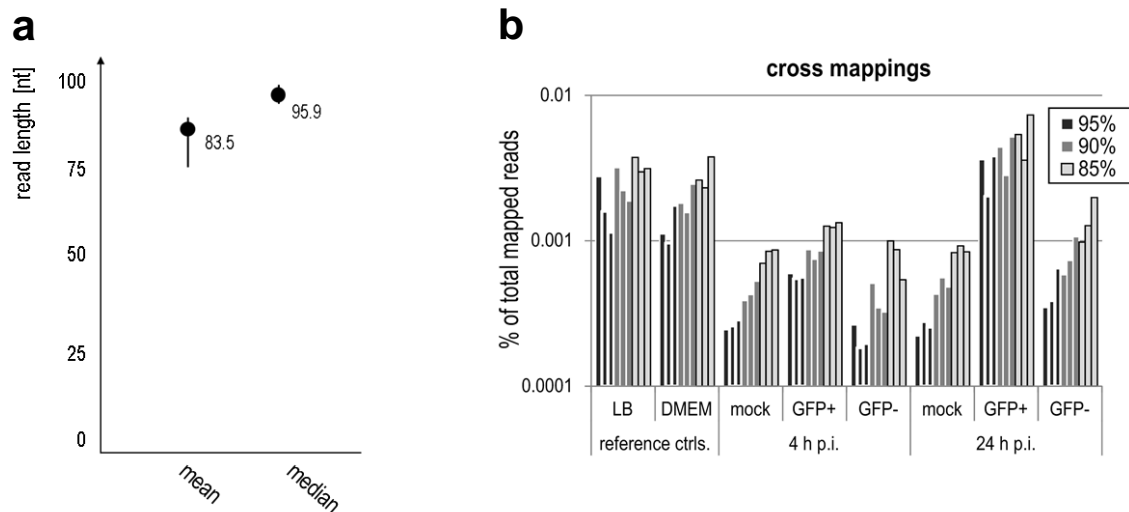


Figure 2.13/Read length distribution and cross-mapping reads. **a**, Mean and median length of raw sequencing reads. **b**, A minor fraction of reads could be mapped to both reference genomes. Their relative amount compared to the total mapped reads is shown for different accuracy cut-offs. Data from the three replicates are shown.

As shown in Fig. 2.14 for representative libraries, by using these settings the discriminatory power of Dual RNA-seq was very high. That is, reads that could be mapped against the *Salmonella* genome (green fractions in Fig. 2.14) were almost exclusively detectable in libraries derived from invaded HeLa-S3 cells while being almost undetected in those derived from non-infected bystander or uninfected control cells. This highlights the stringency of parallel read mapping for the *Salmonella*/human model system. Furthermore, in line with intracellular bacteria undergoing replication over time, a higher proportion of *Salmonella*-derived reads were detected in the 24 h samples compared to the 4 h samples (~2% *Salmonella* reads at 4 h p.i. and ~10% at 24 h p.i.; Fig. 2.14). In particular, between 0.5 M and 3 M *Salmonella*-mapped reads were obtained. With a median read length of 96 nt (Fig. 2.13a), this corresponds to 10-58x genome coverage. The increase in *Salmonella* reads over time matched the relative increase of *Salmonella* RNA in the mixed samples as predicted from qRT-PCR measurements (7 fold increase from 4 h to 24 h p.i.; Fig. 2.8b). On the other hand, the relative fractions of *Salmonella* to human reads in the GFP-positive samples (Fig. 2.14) were slightly higher than predicted (Fig. 2.8b). This suggests either a marginal preference for bacterial over human RNA fragments during cDNA library generation or that the qRT-PCR-based estimation slightly underestimates the relative fraction of *Salmonella* RNA. Nevertheless, even in libraries derived from infected cells (Fig. 2.14) the vast majority (~90-99%; i.e. ~20-35 M reads summing up to ~0.6-1.1x genome coverage) of mapped reads corresponds to the human transcriptome. The fraction of reads that mapped to the spike-in sequence was relatively constant in all samples indicating uniform sequencing (Fig. 2.14).

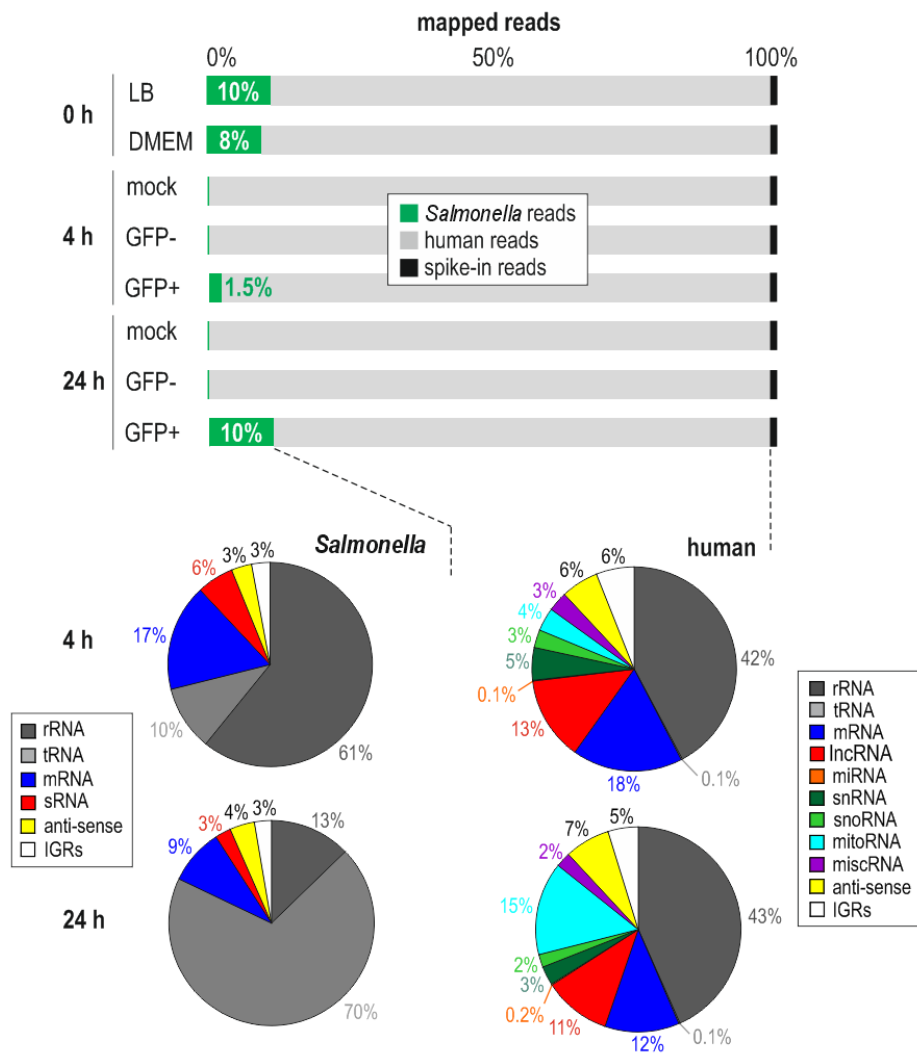


Figure 2.14/Overview of Dual RNA-seq mapping statistics. Upper: Sequencing reads can be stringently assigned to their respective genome or the spike-in sequence. Lower: Dual RNA-seq captures all major RNA classes of bacterial and mammalian organisms. Representative results from one out of the three biological replicates are shown.

Most importantly, all of the major bacterial and human transcript classes were detected in the Dual RNA-seq data (Fig. 2.14, lower part). In the case of the *Salmonella* data subset, the majority of mapped reads (61% in the 4 h sample; dark grey in Fig. 2.14) accounted for ribosomal transcripts. As ribosomal RNA has not been actively depleted, this indicates that in accordance with previous studies (e.g. Sharma et al., 2010) rRNA has a lower conversion efficiency than less structured RNA classes such as mRNAs (9-17%; blue in Fig. 2.14) and sRNAs (3-6%; red), which were readily detectable in the libraries. Strikingly, at 24 h p.i. the vast majority of bacterial reads were mapped to tRNAs, while only a minor proportion were rRNA-derived. This might reflect the reduction in metabolic activity of intracellular *Salmonella* at the late time point (see also Fig. 2.4b). In agreement with previous studies (e.g. Kroger et al., 2012), reads derived from *Salmonella* anti-sense transcripts (yellow) and intergenic regions (white) were lowly abundant.

On the host side, again fewer than expected reads were derived from ribosomal transcripts (~40%; dark grey in Fig. 2.14) and almost no reads mapped to tRNAs. This indicates that the human counterparts of these house-keeping transcripts also have lower conversion rates into cDNA than mRNAs and regulatory non-coding RNA species. Of note, in the present context such under-representation is considered beneficial as it favors sequencing without prior rRNA depletion (as proposed in Westermann et al., 2012). In contrast, more informative human RNA classes such as mRNAs (12-18%; blue in Fig. 2.14), lncRNAs (11-13%; red) and to lesser extent miRNAs (0.1-0.2%; orange) were well covered in the datasets. In conclusion, Dual RNA-seq is able to discriminate between host-pathogen transcriptomes and captures all relevant RNA classes – both coding and non-coding – from two organisms with radically different transcriptome structures.

In order to find the best suited cut-offs for differential gene expression analyses, RPKM values for the individual *Salmonella* and human transcript classes were calculated (Fig. 2.15). On the bacterial side, the average RPKM of mRNAs was ~20 and that of sRNAs >100. Human mRNAs as well as miRNAs were detected with an average RPKM of ~1, while that of lncRNAs was markedly lower, in agreement with previous studies describing that coding genes are generally higher expressed than lncRNA genes (Dunham et al., 2012). The classes of snRNAs and snoRNAs gave rise to an average RPKM of ~0.1 or 10, respectively. Based on these values, we defined the following thresholds for individual RNA classes: For *Salmonella* mRNAs or sRNAs, candidates that exceed a RPKM of 10 or 20, respectively, were considered expressed. For human mRNAs and miRNAs a RPKM cut-off of 0.5 and for lncRNAs a cut-off of 0.1 was set.

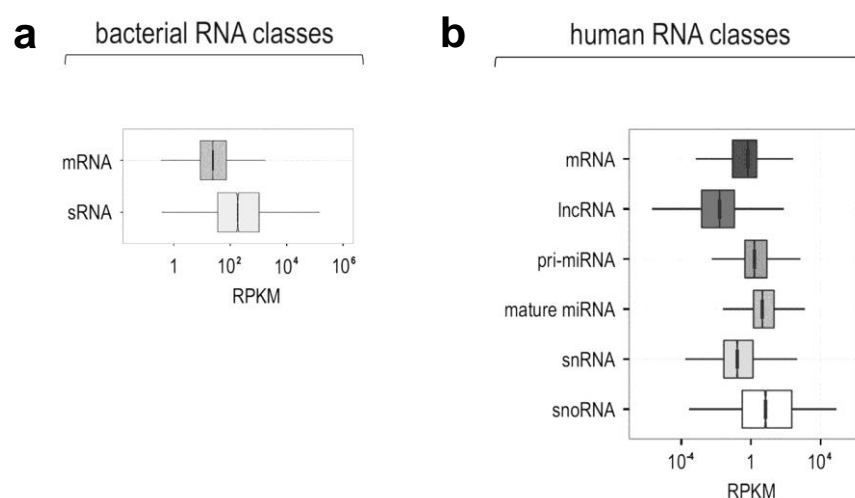


Figure 2.15/RPKM distribution. For the relevant bacterial (**a**) or human (**b**) transcript classes, respectively, RPKM distributions are depicted. The end lines of the box plot indicate the 25 and 75 percent quantiles and the vertical line denotes the median RPKM.

Reproducibility of the Dual RNA-seq method was evaluated by comparing sequencing data between triplicate experiments. Normalized read counts for detected *Salmonella* (mRNA and sRNA) or human transcripts (mRNA, lncRNA, miRNA), respectively, were plotted pair-wise between all biological replicates (Fig. 2.16). Reproducibility of individual transcript detection was found to be high, both for the *Salmonella* ($r = 0.91$ - 0.99) and human datasets ($r = 0.94$ - 0.98). Given this high reproducibility, the data from the three replicates were pooled. In doing so, ~1.5-9 M mapped *Salmonella* reads (30-180x genome coverage) and ~60-100 M mapped human reads (2-3x genome coverage) were obtained.

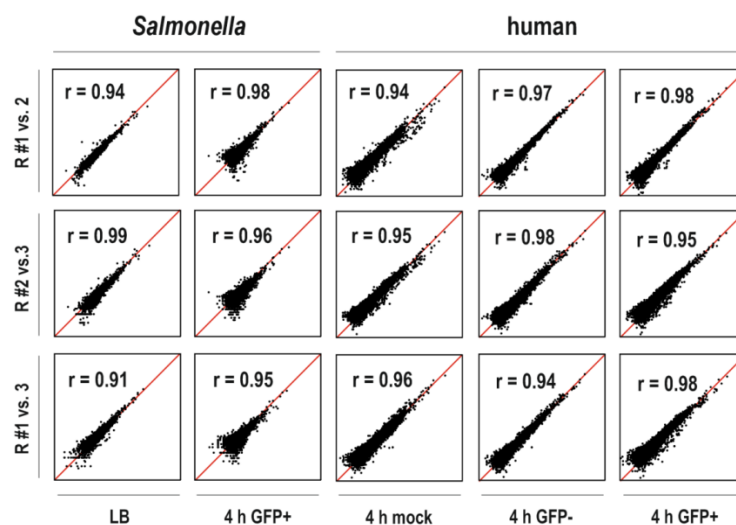


Figure 2.16|Reproducibility of Dual RNA-seq. For the 4 h time point and the respective controls the normalized read counts for each detected transcript (mRNA, lncRNA, miRNA, sRNA) were plotted between biological replicates (log2 scale).

2.7 *Salmonella* gene expression changes revealed by Dual RNA-seq

2.7.1 Impact of culture medium on *Salmonella* gene expression

To perform differential gene expression analyses in intracellular vs. extracellular *Salmonella* we first sought a relevant reference control. In particular, RNA was collected from extracellular bacteria either in rich LB medium or in host cell DMEM (Fig. 2.11). When comparing gene expression levels in the two media substantial differences were observed (Fig. 2.17). For instance, according to previous studies (e.g. Hautefort et al., 2008) iron transporter genes (e.g. *fepD* (Tsolis et al., 1995)) were strongly activated upon medium transfer to DMEM as were the cold-shock-induced genes *lpxP* (Carty et al., 1999) and *cspA* (Jeffreys et al., 1998). Likewise, expression levels of sRNAs were strongly affected by the medium shift. Note however, that while several sRNAs that were subsequently identified to be up-regulated inside the host cell (MicL,

RyhB, IsrE, RybB) were already induced in DMEM itself, STnc440 was in fact down-regulated indicating that its high induction upon host cell entry (see below) is not related to the infection medium. Together, based on these findings extracellular *Salmonella* in DMEM were chosen as reference control for all further comparisons. If the LB sample would be used instead, one would have to dissect two overlaying gene expression programs – one triggered in response to metal ion limitation in the host cell medium and the one to the adaptation to an intracellular lifestyle.

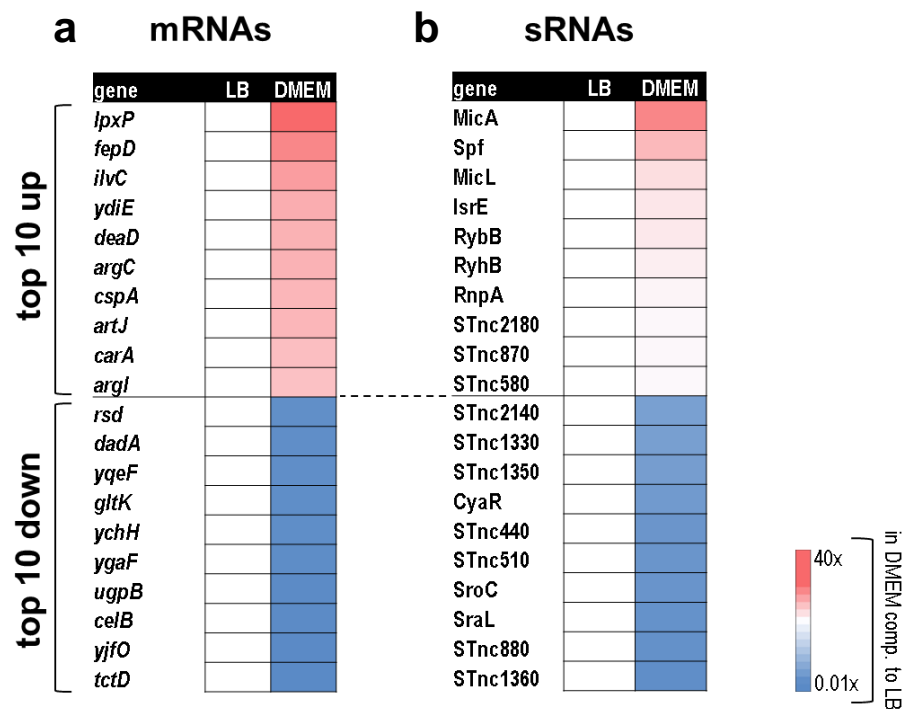


Figure 2.17|Influence of the medium shift on *Salmonella* gene expression. *Salmonella* was grown in LB to an OD₆₀₀ of 2 and then transferred to DMEM. RNA-seq was performed on samples either before or after the transfer. The ten most strongly induced or repressed mRNAs (a) or sRNAs (b) in DMEM compared to LB are shown. Biological triplicates were considered.

In vivo, *Salmonella* invades two major cell types, epithelial cells and monocytes/macrophages. To assess the differences and similarities in *Salmonella* expression programs triggered during the infection of these fundamentally different cell types, the analysis was extended to a second human model cell line. In addition to the infection of epithelial HeLa-S3 cells (biological triplicates) we included samples taken 4 h or 24 h after infection of monocyte-like, non-differentiated THP-1 cells (single replicate) (Fig. 2.18).

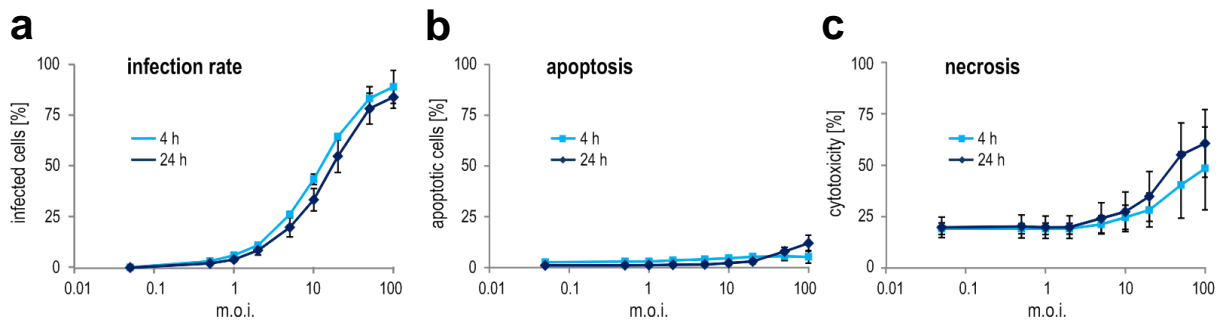


Figure 2.18/Pathogen dose effects on infectivity and cell death of human THP-1. Analogous to Fig. 2.3. For all further experiments with THP-1 cells, a m.o.i. of 10 was used for infection.

2.7.2 Virulence gene expression during infection

During infection of HeLa-S3 cells *Salmonella* show a biphasic pattern of virulence gene expression (Schlumberger and Hardt, 2006): As invasion of non-phagocytic cell types such as epithelial HeLa-S3 is T3SS-1-mediated, SPI-1 genes should be induced early during the process, i.e. already in extracellular *Salmonella*. For example, previous work reports SPI-1 activation when *Salmonella* were grown to late exponential/early stationary phase *in vitro* (Kroger et al., 2012, 2013). In contrast, the T3SS-2 is required to establish the intracellular lifestyle and consequently SPI-2 genes are activated only after host cell invasion at a time when SPI-1 is already repressed to basal levels. Here, relative expression levels of SPI genes were monitored at the two time points after HeLa-S3 invasion (Fig. 2.19a). Indeed, while SPI-2 genes were strongly induced inside the host cell niche at both time points, SPI-1 genes were down-regulated compared to their extracellular levels, especially at 24 h p.i. Interestingly, similar expression patterns of SPI genes were found after infection of monocytic THP-1 cells (Fig. 2.19b): Again SPI-2 genes were generally induced post-invasion and SPI-1 genes suppressed. However, full deactivation of SPI-1 was already observed at the early time point, suggesting either a fast shut-off of invasion genes after host cell internalization or that *Salmonella* may invade this cell type in a T3SS-1-independent manner. A second major difference between the two infections concerned the *avrA* gene, whose expression was rapidly down-regulated upon invasion of HeLa-S3 but induced upon THP-1 invasion. Since *avrA* encodes a translocated effector acetyltransferase with immune-suppressive functions (Hardt and Galan, 1997), its differential expression between epithelial and monocytic cells might reflect distinct requirements for interfering with cell type-specific host defense mechanisms.

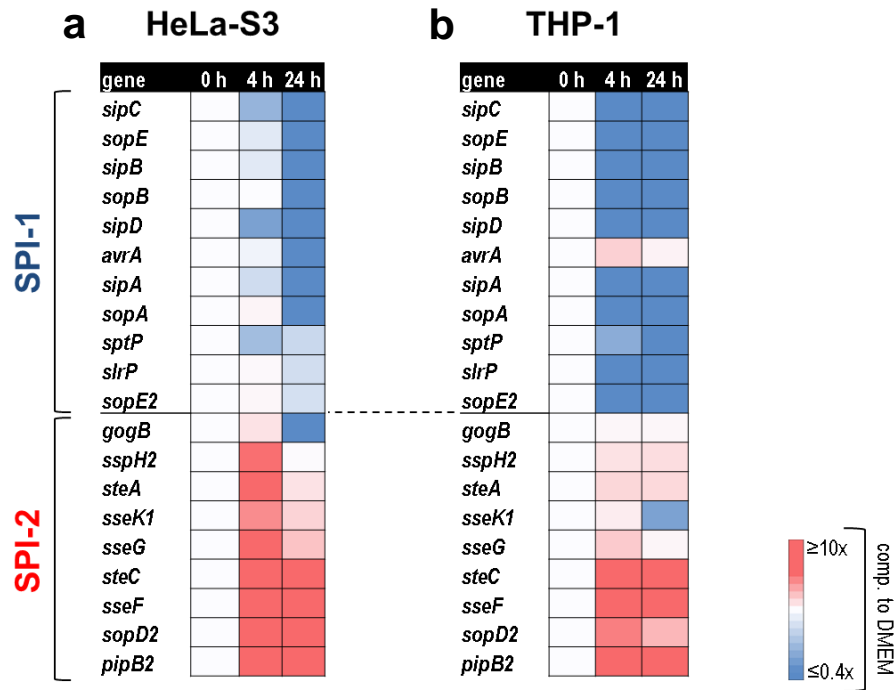


Figure 2.19 *Salmonella* virulence gene expression patterns in different host cell types. Relative gene expression changes of the detected SPI-1 or SPI-2 genes at different time points after infection of HeLa-S3 (a) or THP-1 cells (b) compared to the extracellular reference control ('DMEM').

Three further pathogenicity islands of *S. Typhimurium* (SPI-3/-4/-5) are known to be involved in virulence (Fabrega and Vila, 2013). However, these have been studied in much less detail than SPI-1 and -2. For example, only little is known about their intracellular expression status. Here inside both cell types SPI-5 seemed to follow similar kinetics as SPI-2 (Fig. 2.20), i.e. genes from both islands were intracellularly induced. In agreement with previous findings that reported a co-regulation of SPI-4 with invasion genes (Gerlach et al., 2007), SPI-4 expression was rapidly shut off after invasion, thereby resembling the expression kinetics of SPI-1 factors. Genes from SPI-3 displayed a bipartite expression pattern: The *mgt* operon was strongly activated inside both cell types according to previous reports (BlancPotard and Groisman, 1997; Hautefort et al., 2008), whereas other SPI-3 genes were repressed. In summary, characteristic expression signatures of virulence genes fit the dogma of SPI-1 genes being required for host cell invasion and SPI-2 genes for intracellular survival and replication, thereby providing proof-of-principle for Dual RNA-seq to reliably capture *Salmonella*'s intracellular gene expression changes. Moreover, the approach yielded insights into the intracellular expression levels of the under-studied SPI-3, -4 and -5 loci.

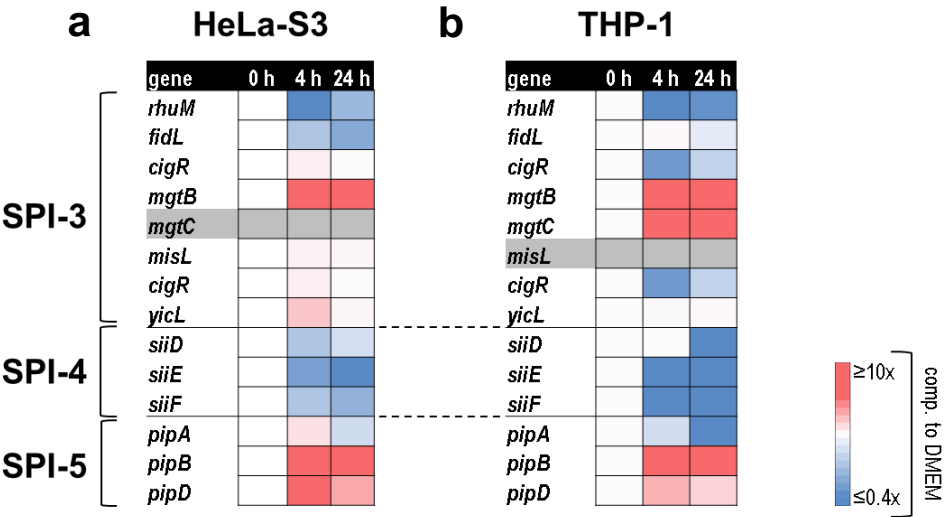


Figure 2.20|Expression of genes from pathogenicity islands SPI-3 to SPI-5. Analogous to Fig. 2.19. Grey shaded boxes mean that the respective transcript was not detected in at least one of the three time points.

Another hallmark of bacterial invasion is the frequent modification of cell wall LPS to counter-act PAMP sensing by host PRRs (e.g. Guo et al., 1997; Matsuura et al., 2012; Paciello et al., 2013). This occurs because in such cases the intracellular LPS form is hardly sensed by host receptors and thus the modification facilitates the evasion of the host’s immune response and becomes important predominantly in the context of chronic infections. Besides LPS, lipoproteins represent a second integral component of the outer membrane of Gram-negative bacteria. *Salmonella* carries two functional copies of the lipoprotein gene in its chromosome (*lppA* and *lppB*). Interestingly, Dual RNA-seq reports a switch from *lppA* to *lppB* expression during the course of infection of HeLa-S3 cells (Fig. 2.21). It is therefore tempting to speculate that in addition to the modification of the LPS component lipid A (Guo et al., 1997; Matsuura et al., 2012), *Salmonella* has evolved mechanisms to ensure contextual expression of its different lipoprotein variants during infection what might further contribute to the evasion from host recognition. Note that due to the lower sequencing depth, *lpp* expression was not detected in THP-1 infections.

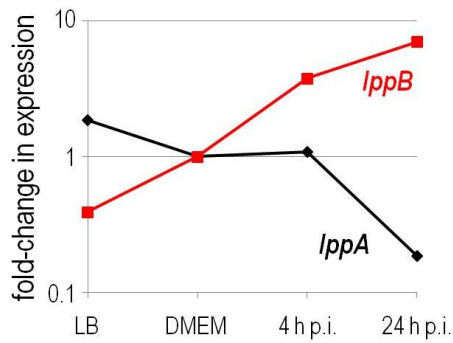


Figure 2.21|A switch in *lpp* expression upon HeLa-S3 invasion. Differential expression of *lppA* (black) and *lppB* (red) mRNAs over the indicated conditions (relative to the DMEM reference control).

2.7.3 *Salmonella* encounters various stresses inside its host cell niche, which are reflected in its gene expression profile

The obtained Dual RNA-seq data can further provide information on the different types of stresses *Salmonella* faces within its host cell niche. For instance, after infection of both HeLa-S3 and THP-1 cells, changes in the expression of genes involved in the response to iron (e.g. induction of *iro* mRNAs as well as RyhB and IsrE sRNAs (Masse and Gottesman, 2002)) and magnesium limitation (*mgtB*) were observed (Fig. 2.22a). This correlates with previous findings (Hautefort et al., 2008; Ortega et al., 2012) and indicates that iron and magnesium deficiency are generic sources of stress that intracellular *Salmonella* has to cope with irrespective of the host cell type.

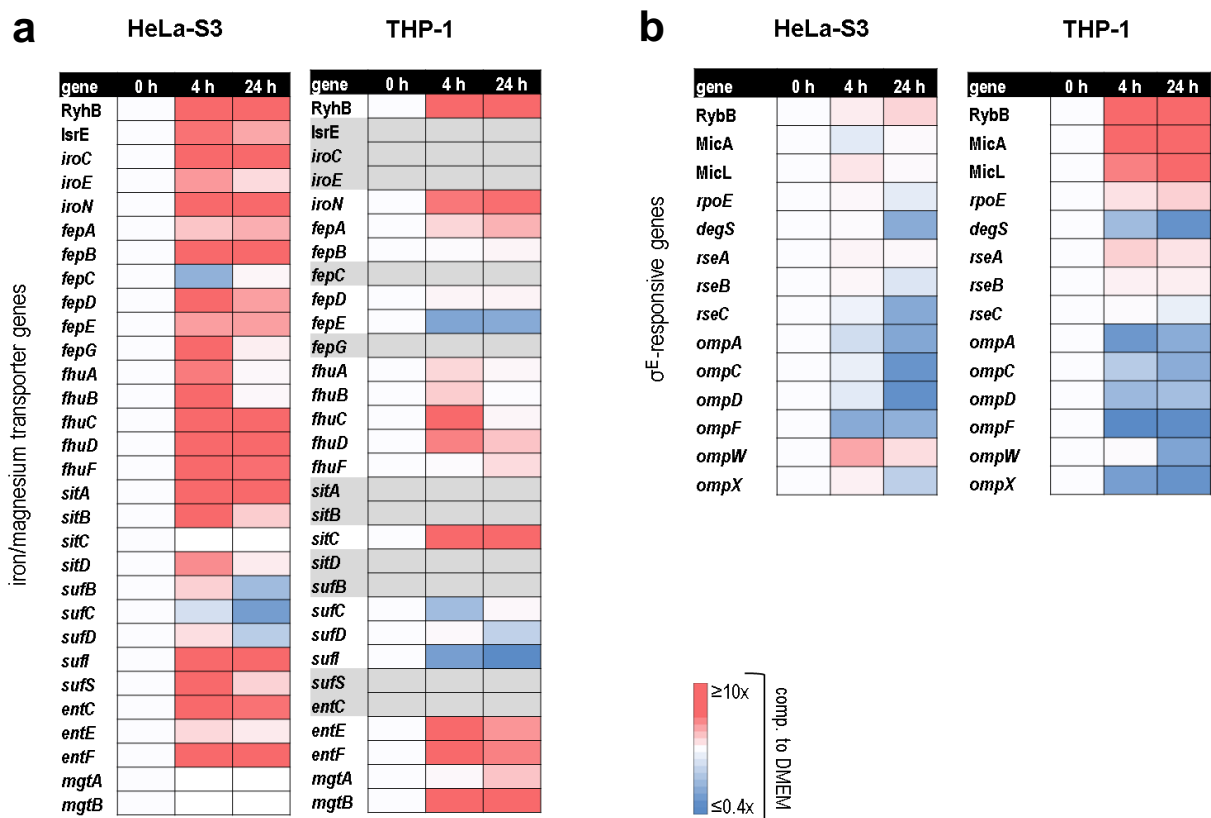


Figure 2.22|Metal ion scarcity and membrane stress are reflected in the transcriptome of intracellular *Salmonella*. **a**, Expression of genes involved in iron or magnesium uptake. Undetected genes in the THP-1 infection experiment are grey-shaded. **b**, Regulation of genes implicated in the σ^E response to cell surface stress.

Besides metal ion deficiency, hallmarks of the envelope stress response were observed in the Dual RNA-seq data – predominantly inside monocytic cells (Fig. 2.22b): For example, the σ^E -encoding *rpoE* mRNA itself, the σ^E -dependent *rse* operon and σ^E -dependent sRNAs such as RybB,

MicA and MicL (Guo et al., 2014; Hobbs et al., 2010; Papenfort et al., 2006) were induced upon THP-1 invasion. At the same time numerous *omp* mRNAs, which are major targets of σ^E -activated sRNAs (Rhodius et al., 2006), were rapidly down-regulated. It is tempting to speculate that activation of the σ^E response and the resulting suppression of multiple outer membrane proteins might represent another immune evasion strategy of *Salmonella*. Moreover, this type of adaptation seems predominantly important inside monocytes, which express a wide range of cytosolic PRRs (Takeshita et al., 2004; Zhao et al., 2007).

At the late time point of infection, invading *Salmonella* no longer proliferated (Fig. 2.4b). This was reflected in the composition of their transcriptome (Fig. 2.14). For instance, 24 h after HeLa-S3 infection (but not that of THP-1; not shown) a massive increase in the fraction of reads mapped to tRNAs was observed, whereas rRNAs were lowly abundant (Fig. 2.14). Closer inspection of the most abundantly detected transcripts revealed the three initiator tRNAs-fMet (tRNA051, -052 and -056) to be the source of the increase. The coverage plots of these RNAs revealed a sharp decline in the profiles near the center of the annotated tRNA gene (indicated by red arrows in Fig. 2.23a), pointing at a processing event. Interestingly, cleavage was found to be specific to initiator tRNAs since elongator tRNAs were not affected as validated by Northern blot (Fig. 2.23b). tRNA-fMet processing seems not to be specific to *Salmonella* growing inside HeLa-S3. Rather extracellular *Salmonella* growing to stationary phase in LB likewise displayed tRNA cleavage (Fig. 2.23c), suggesting it might be a generic feature of bacteria reaching a certain cell density. In addition, cleavage was also observed for *Salmonella* that had been treated for 4 h with high concentrations of gentamicin (Fig. 2.23d). As this antibiotic acts by targeting the 30S subunit of bacterial ribosomes to prevent mRNA translation, the observed effect might indicate that at the late infection time point traces of the antibiotic reached the host cell cytosol.

Initiator tRNA cleavage has previously been described in *Salmonella* and shown to be in part mediated via the toxin/anti-toxin system (TA system) VapB/C (Winther and Gerdes, 2011). Here, while the abundance of *vapC* mRNA increased over time (Fig. 2.23e), in the present case deletion of *vapB/C* from the *Salmonella* genome did not significantly affect tRNA-fMet cleaving (Fig. 2.23f, g). Together this indicates that initiator tRNA processing in intracellular *Salmonella* during the late infection stages might be caused by diverse bacterial stresses, such as a high cell density or gentamicin traces reaching the host cell's cytosol. However, as processing was still seen for the respective deletion mutant, VapB/C is unlikely to be the major player in this context.

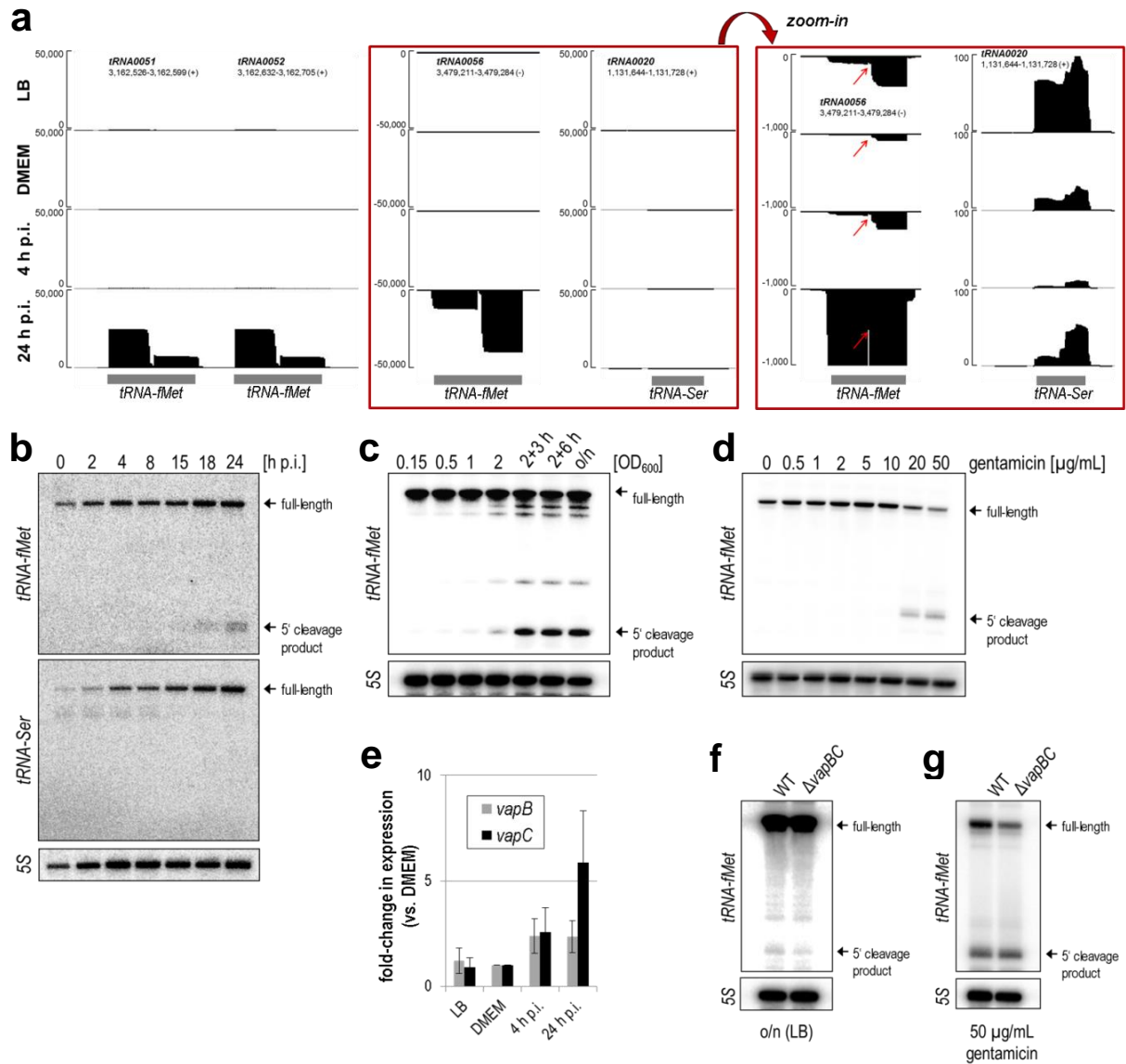


Figure 2.23|High bacterial cell densities and gentamicin concentrations are accompanied by initiator tRNA cleavage. **a**, Coverage plots for individual tRNA loci in the *Salmonella* genome. tRNA0051, -52 and -56 encode initiator tRNA (tRNA-fMet) and tRNA0020 is an elongator tRNA (tRNA-Ser). Transcript annotations are depicted below the coverage plots. Red box: Magnification of one of the tRNA-fMet and the tRNA-Ser loci. Red arrows indicate the position of a cleavage event for tRNA-fMet. **b**, Northern blot analysis of RNA samples isolated from *Salmonella*-infected HeLa-S3 cells at different time points using a radio-labeled DNA probe against the 5' end of initiator tRNA-fMet or elongator tRNA-Ser. *Salmonella* 5S rRNA serves as loading control. **c**, Analogous Northern analysis of RNA samples from in vitro-grown *Salmonella* (in LB). **d**, In vitro-growing *Salmonella* were treated for 4 h with gentamicin in the indicated final concentrations prior to RNA isolation and Northern analysis as before. **e**, Dual RNA-seq data reveals an induction of *Salmonella* vapC mRNA at 24 h p.i. Shown is the mean \pm SD from the triplicate experiment. **f**, **g**, Wild-type *Salmonella* (WT) or a vapB/C deletion mutant were grown overnight in LB medium (**f**) or treated for 4 h with 50 μ g/mL of gentamicin (**g**) prior to RNA extraction and Northern blot analysis.

2.7.4 sRNA expression profile of intracellular *Salmonella*

Many sRNAs were identified to be strongly regulated during infection (Fig. 2.24a,b). To confirm Dual RNA-seq-based gene expression profiling of intracellular *Salmonella* by unrelated methods, four of the induced sRNAs were selected for independent qRT-PCR experiments: The well-characterized *Salmonella* sRNAs RyhB (Masse and Gottesman, 2002) and RybB (Balbontin et al., 2010; Papenfort et al., 2006) were chosen as representative for a highly (>100 fold) or slightly (~2 fold) induced sRNA. In addition, STnc440 – a previously validated (Sittka et al., 2008) but poorly characterized sRNA – was included as was a computationally predicted candidate sRNA, STnc510. The latter two showed very similar expression patterns with ~100-150 fold induction at 4 h p.i. in HeLa-S3 and ~30-50 fold induction in THP-1. These sRNAs were well covered with sequencing reads (Fig. 2.24c), and qRT-PCR measurements validated their observed up-regulation (Fig. 2.24d). Importantly, the approximate fold-changes of induction were very similar among the two methods (Fig. 2.24d). In the qRT-PCR data, STnc440 was induced by ~139 fold at 4 h p.i. (compared to ~143 fold in Dual RNA-seq) and RybB by ~5 fold (compared to 2 fold). One of the largest discrepancies between the two methods (39 vs. 119 fold) was observed for RyhB. This over-estimation by deep-sequencing might be due to the low abundance of the RyhB sRNA in the extracellular control (i.e. only 22 or 50 cumulated reads for the replicate LB or DMEM libraries, respectively) (Fig. 2.24c). If this would be the reason then such obstacles may be solved simply by increasing the sequencing depth. Indeed, when sequencing deeper the resulting fold-change in RyhB induction was similar to that deduced from qRT-PCR experiments (see the third chapter of this study).

Besides these validated sRNAs, bioinformatically predicted sRNA candidates are also included in our annotation (Förstner et al., unpublished). The latter, however, have not yet been manually confirmed as *bona fide* sRNAs. One such candidate that was strongly induced after invasion was STnc510 (Fig. 2.24). However, closer inspection of STnc510 indicated that it may represent a non-annotated 5' UTR of the downstream *pagC* coding gene rather than a true sRNA (Fig. 2.24c). In fact, STnc510 was always co-expressed with *pagC* and with a size of 810 nt it would be atypically long for a *bona fide* sRNA. Thus, STnc510 was discarded from the list of *Salmonella* sRNAs for the following analyses.

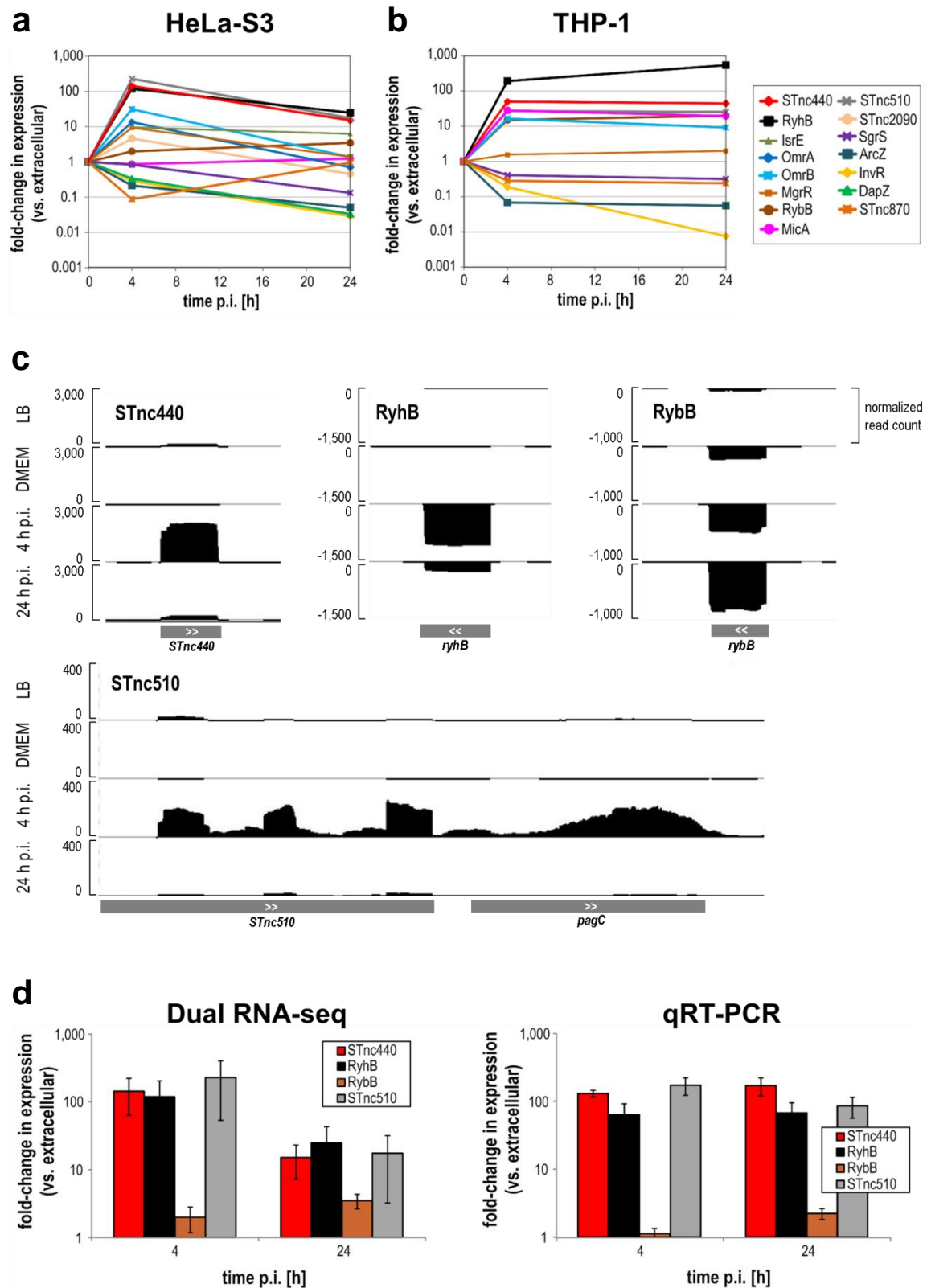


Figure 2.24|Intracellular sRNA expression profiles. *a,b*, Dual RNA-seq-based sRNA expression changes during infection of HeLa-S3 (*a*) or THP-1 cells (*b*). Depicted are sRNAs which are referred to in the main text. Note that not all of the sRNAs depicted in panel *a* were detected inside THP-1 cells and thus, some genes are missing in panel *b*. *c*, Coverage plots for the four selected sRNAs. Annotated transcripts are shown at the bottom of the coverage plots. Depicted are the read distributions from a representative replicate experiment. *d*, Independent validation of Dual RNA-seq results (left panel) by qRT-PCR (right) for the expression of the same four sRNAs inside HeLa-S3. The data represent the mean \pm SD from biological triplicates.

2.7.5 Global overlap between *Salmonella* gene expression inside epithelial and monocytic cells

The above described results revealed several specific differences between *Salmonella* gene expression inside epithelial and monocytic cells. For instance, the opposing expression of the virulence gene *avrA* (Fig. 2.19), the exclusive tRNA cleavage inside HeLa-S3 (Fig. 2.23) or a full σ^E response only within THP-1 (Fig. 2.22b) are prominent host cell type-specific differences. Overall, however, there seemed to be a surprisingly huge overlap in *Salmonella* gene expression profiles between the infections of epithelial and monocytic cells. To address whether these similarities reach beyond the manually inspected cases, a global overlap between differentially (≥ 2 fold) regulated mRNAs or sRNAs after HeLa-S3 or THP-1 infection was determined (Fig. 2.25). This revealed a slightly greater overlap in sRNA (38-55%) than in mRNA expression (15-49%). Interestingly, gene expression of intra-epithelial and intra-monocytic *Salmonella* was more similar at the later time point than early after invasion, and this was true both for mRNAs (4 h: 15%; 24 h: 33-49%) and sRNAs (4h: 38-41%; 24h: 55%). This may indicate that more specific gene sets are expressed to initially adapt to the respective host environment, whereas during the late infectious stages relatively similar expression programs are elicited in either cell type. In general, however, the commonalities between *Salmonella* inside epithelial and monocytic cells indicate that a core set of genes is required to establish and maintain the intracellular lifestyle.

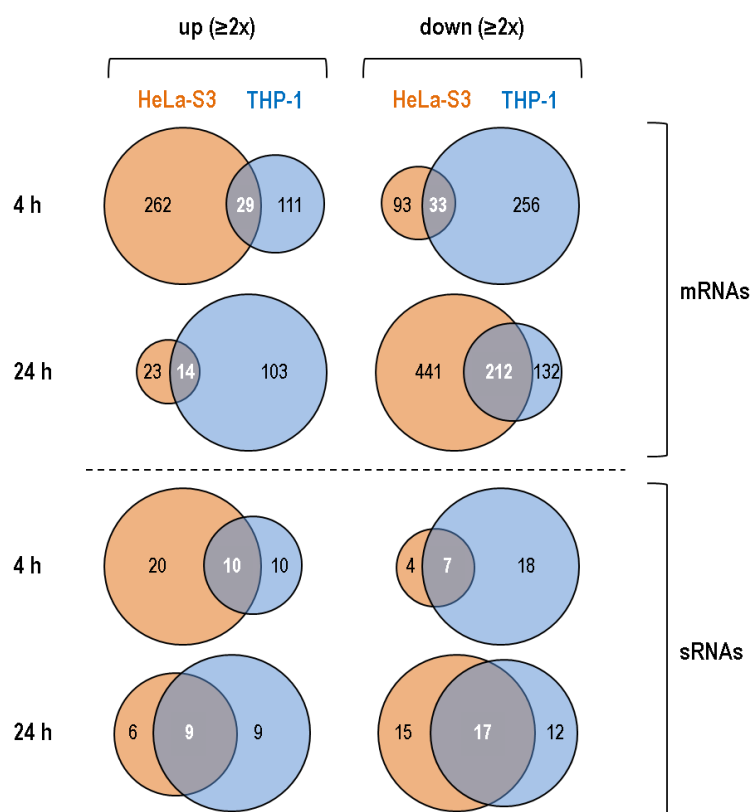


Figure 2.25/Overlap in *Salmonella* gene expression between the infection of epithelial and monocytic cells. Shown are all detected mRNAs and sRNAs that were regulated (≥ 2 fold) under at least one condition.

In summary, using the Dual RNA-seq approach previously reported gene expression programs could readily be recapitulated. In addition, global communalities as well as host cell type-specific changes in *Salmonella* gene expression were revealed. This broadens our understanding of the diverse types of stresses *Salmonella* faces when entering its host cell environment. Finally, several previously under-characterized sRNAs were found to be regulated during infection – a striking example being STnc440. Its strong intracellular induction inside both cell types suggests a general role for this sRNA in virulence (see chapter 3).

2.8 Changes in the host transcriptome

2.8.1 Comparison of invaded host cells and non-infected bystanders

Due to the low sequencing depth and lack of replicates for the THP-1 infection, analysis of the host response was restricted to HeLa-S3. In the following, the main findings from the pilot experiment about the host response will be described. Importantly, all major human RNA classes were detected by Dual RNA-seq (Fig. 2.14) allowing for the calling of differentially expressed genes from all relevant classes. Due to the overall subtler expression changes in the host, a different cut-off for differential expression was set (≥ 1.5 fold) than had been used for the pathogen (≥ 2 fold).

Global expression profiles were compared between invaded and bystander cells. This unveiled substantial differences in the expression programs, e.g. that of mRNAs (Fig. 2.26), between the two sub-populations. As expected, generally more changes in gene expression were observed in the fraction of GFP-positive host cells that carry replicating pathogens than in non-infected bystander cells. Although the dataset cannot discriminate between changes induced following bacterial sensing by host PRRs and those elicited by the active manipulation of the host by the pathogen, it suggests the two host sub-populations to have fundamentally different physiological states. This highlights the importance of the FACS-based separation step.

Since the infection of epithelial cells with *Salmonella* represents a well-studied system, at least for the expression of coding genes, existing microarray datasets are available and compared with the host subset of the Dual RNA-seq data. There was an overall good agreement between these datasets, e.g. the two most strongly induced mRNAs in infected cells at 4 h – *SERPINB3* and *Coagulation factor II receptor (F2R)* – (Fig. 2.27) were found to be up-regulated also in previous studies (e.g. Bruno et al., 2009; Hannemann et al., 2013). Other early induced mRNAs such as *IL6* and *TNFAIP3* (Fig. 2.27) have been defined as an integral part of the “common host response” of mammalian cells to diverse pathogen species (Jenner and Young, 2005). Strikingly, expression data indicated a general transcriptional shut-down in invaded host cells at 24 h. For example, of

the 13,126 detected mRNAs, 7,858 (60%) were down-regulated (≥ 1.5 fold) under these conditions (Fig. 2.26).

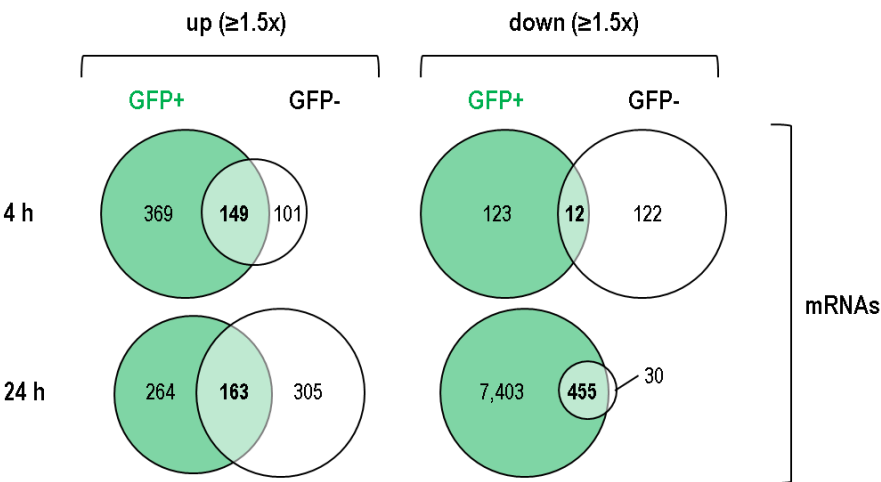


Figure 2.26/Overlap of differentially regulated human mRNAs between invaded and bystander cells. Shown are all detected mRNAs that were regulated (≥ 1.5 fold) under at least one condition.

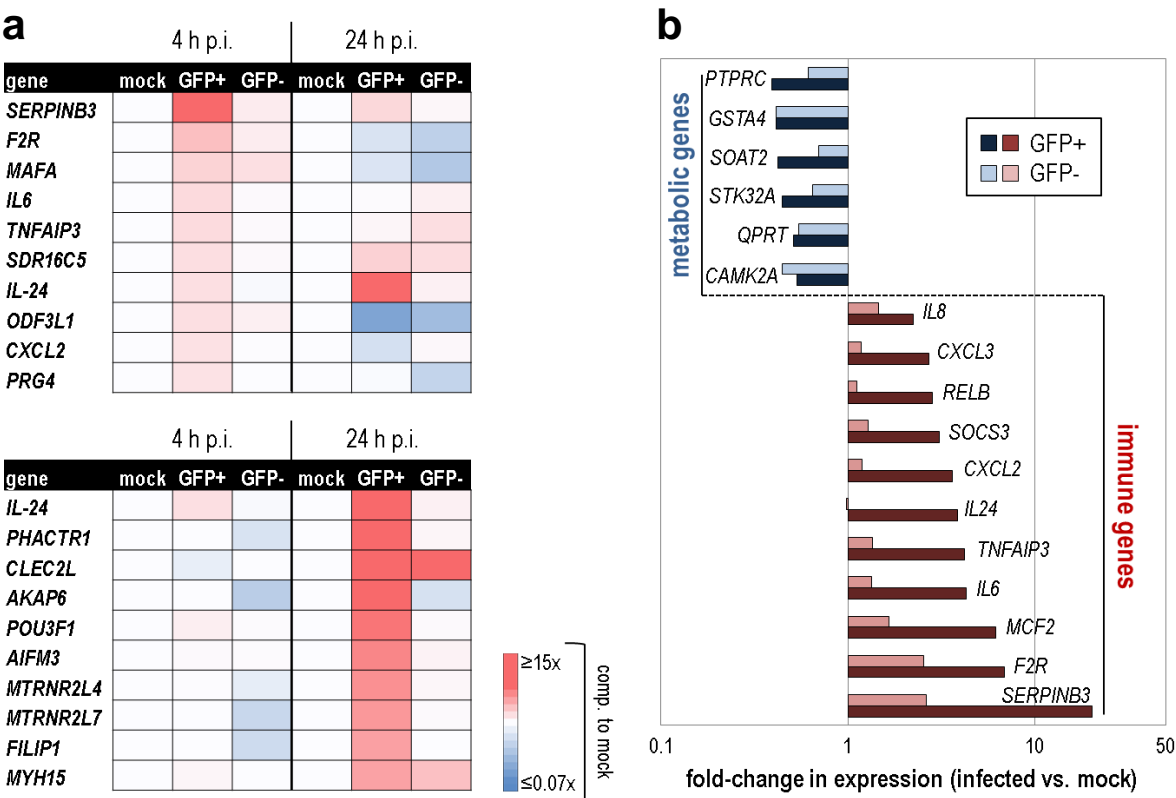


Figure 2.27/Immune genes are strongly induced after invasive infection. **a**, The ten most strongly induced genes in invaded HeLa-S3 cells at 4 h (upper panel) or 24 h p.i. (lower panel) are depicted. **b**, Several genes involved in metabolic pathways are down-regulated at 4 h p.i. in both invaded and bystander cells. At the same time many immune genes are strongly up-regulated, predominantly in invaded cells. All data reflect biological triplicates.

2.8.2 Expression changes of NFκB-responsive and mitochondrial genes

Among the common features displayed by pathogen-infected mammalian cells is the decreased expression of many metabolic genes and an activation of immune factors. For instance, a recent study described a general repression of metabolic genes in cells of the murine colon upon *S. Typhimurium* infection (Liu et al., 2010). In the present study, a down-regulation of several metabolic enzymes was observed at 4 h p.i. in both host cell sub-populations (Fig. 2.27b). In addition, certain immune-associated genes including *IL6*, *IL8* and *IL24* as well as *chemokine ligand 2 (CXCL2)* and *TNFAIP3* were rapidly induced in invaded HeLa cells at 4 h p.i. (Fig. 2.27b).

Cell surface receptors such as the LPS-sensing TLR4 or bacterial lipoprotein-responsive TLR2 are involved in the detection of extracellular pathogens. Thus, due to the lack of functional TLR2 and TLR4 signaling (Wyllie et al., 2000), HeLa cells are commonly considered unresponsive to external PAMP stimulation. Indeed, the non-infected bystander cells did not trigger an obvious immune response. It appears thus likely, that the innate anti-microbial response in HeLa cells is primarily mediated by intracellular stimuli, e.g. through NOD1 which is expressed in HeLa cells (Buchholz and Stephens, 2008). As non-infected bystanders, however, represent by far the major population in the culture (≥95%; Fig. 2.3a) inflammatory responses triggered by the few invaded cells would be difficult to detect without the FACS enrichment step.

The immediate-early response transcription factor NFκB is a central activator of the innate immune response (Naumann, 2000). Bacterial endotoxins as well as cytokine stimulation can activate cellular signal transduction pathways that converge at the phosphorylation and nuclear import of NFκB. Once in the nucleus, NFκB initiates a global response by regulating the transcription of its target genes which often encode cytokines, anti-apoptotic proteins, or additional transcription factors (Oeckinghaus and Ghosh, 2009). Since HeLa-S3 cells were found to be non-responsive to extracellular PAMP stimulation (see above) early NFκB activation can be expected to be an exclusive feature of invaded cells. Indeed, as shown in Figure 2.28a at 4 h p.i. a set of known NFκB target genes were selectively activated in GFP-positive but not -negative cells. At the late stage of infection, however, NFκB-responsive genes such as *IL8* or *TNFAIP3* were also induced in bystanders. This can be explained by the release of inflammatory cytokines by invaded cells that would spread in the culture and eventually elicit activation of the NFκB cascade also in non-infected neighboring cells.

The unbiased nature of Dual RNA-seq not only enables the quantification of expression of nuclear but also organellar genes. At the late stage of infection mitochondrial transcripts were abundantly detected in invaded hosts (Fig. 2.14 and 2.28b). While the underlying molecular mechanisms remain to be determined, this finding might place *Salmonella* on the emerging list of intracellular pathogens that alter mitochondrial activity of their host (Rudel et al., 2010).

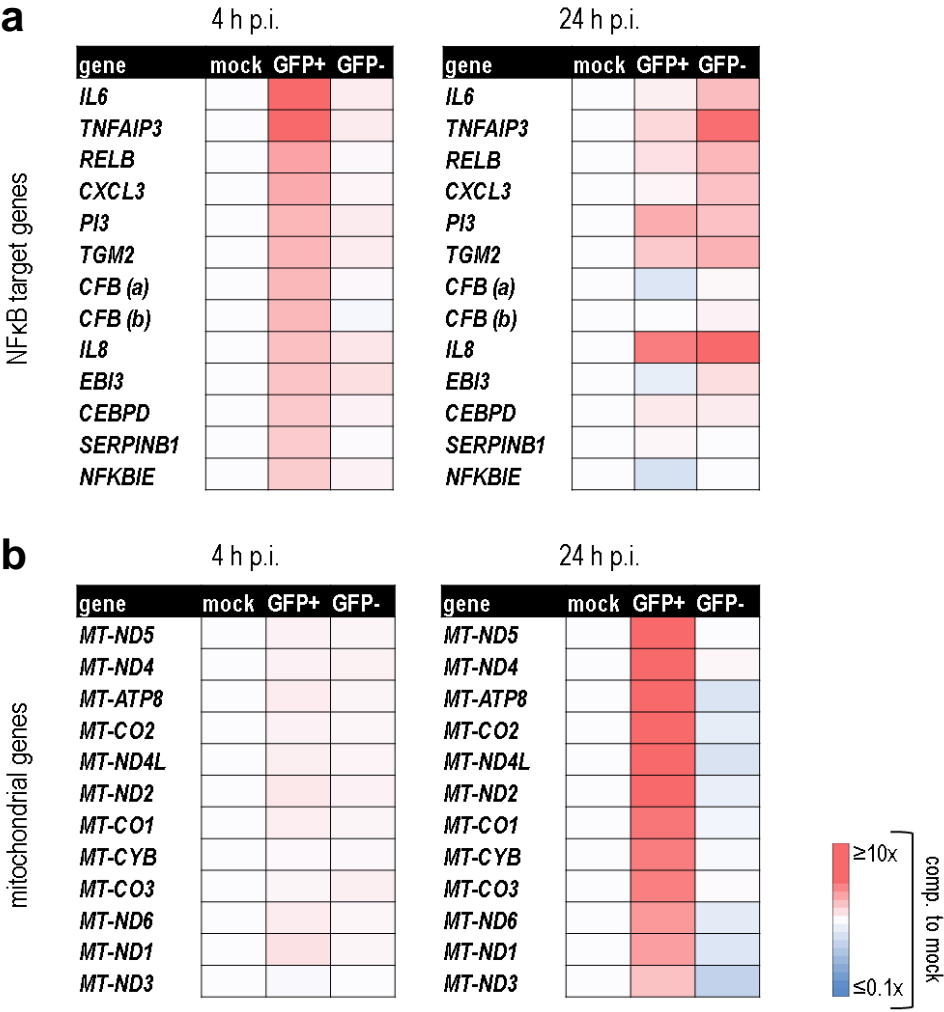


Figure 2.28 *NFκB targets and mitochondrial genes are induced in invaded HeLa-S3 cells. a*, At 4 h p.i. NFκB activation was limited to invaded host cells, whereas at 24 h p.i. NFκB-responsive genes were induced also in non-infected bystanders. **b**, Mitochondrial mRNAs accumulated at 24 h p.i. in invaded HeLa-S3 cells.

2.8.3 The non-coding response

This study provides the first global map of both polyadenylated and non-polyadenylated lncRNAs that respond to bacterial infection. Typically, a properly set RPKM-based selection filter allows for the discrimination between reliably detected transcripts and background noise. As lncRNAs, however, only recently emerged as a major subject of research their annotations are sometimes of poor quality and might inflate the RPKM calculation. As a consequence, for lncRNAs it is eminently important to inspect the read distribution along the annotated transcript to check for a good coverage. For instance, two lncRNA candidates that were both differentially regulated and well-covered with reads in the IGB profiles were *uc002pec.2* and *RP11-255B23.3* (Fig. 2.29a). The read distribution for the former revealed the presence of a 14 kb-long RNA and thus matched well to the annotated transcript structure. In contrast reads exceeded the annotation of *RP11-255B23.3* at both ends, suggesting the existence of a continuous transcript of

~120 kb. Both lncRNA candidates were identified as being repressed during the late infection stage. However, while *RP11-255B23.3* was down-regulated specifically in invaded cells and might thus reflect the general transcriptional shut-down observed under that condition (see above), *uc002pec.2* expression was suppressed in both sub-populations. qRT-PCR validation measurements correlated well with the Dual RNA-seq-based expression values (Fig. 2.29b). Mechanistic and functional implications of these lncRNA candidates were not further addressed. However, the data highlights that Dual RNA-seq profiling is not restricted to mRNAs but captures also non-coding expression changes of the host during bacterial infection.

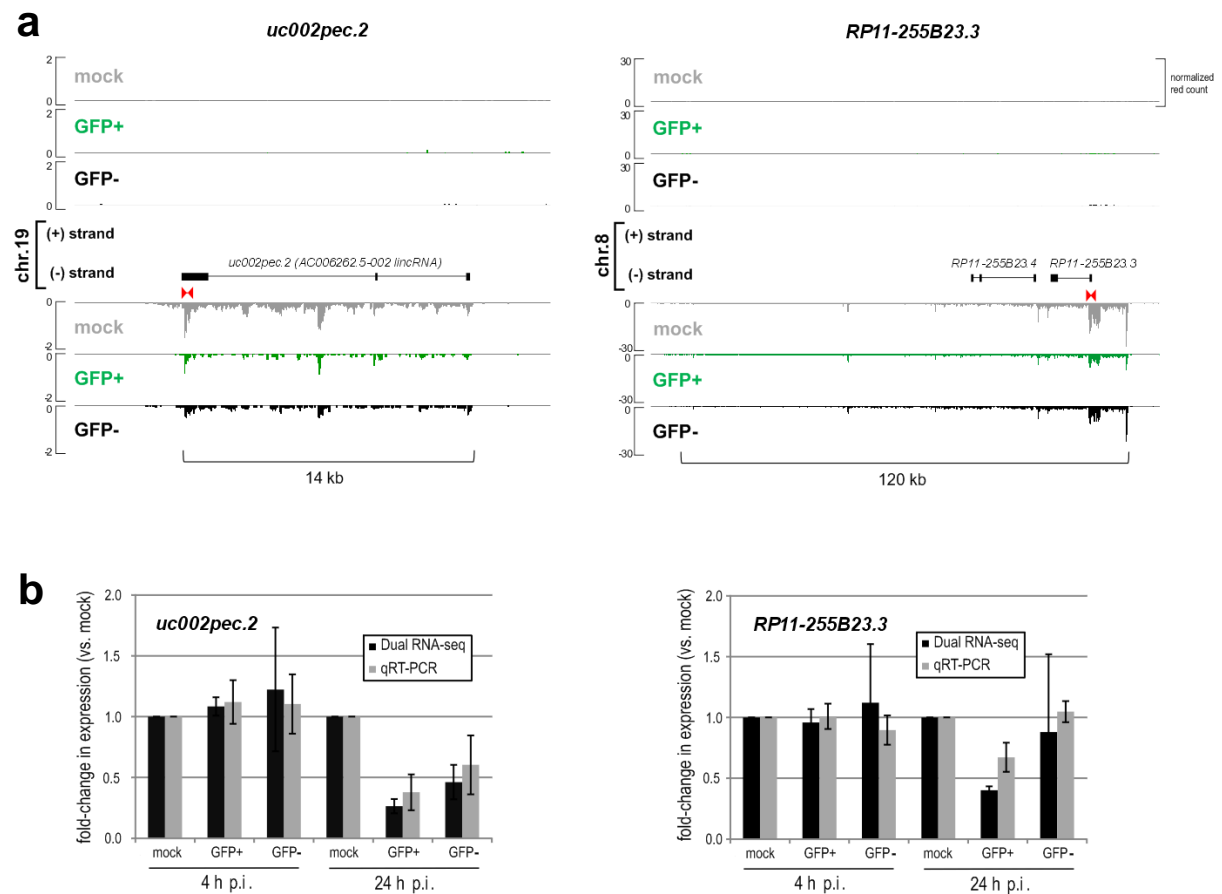


Figure 2.29|Differential expression of human lncRNAs after Salmonella infection. **a**, Coverage plots for two human lncRNAs that were down-regulated after 24 h of infection. Transcript annotations are indicated. Shown are the read distributions from a representative replicate experiment. **b**, Confirmation of the regulation by qRT-PCR. Mean \pm SD from biological triplicates are shown.

Besides lncRNAs, also miRNAs can be detected using the Dual RNA-seq approach, albeit with only a minor read fraction (see Fig. 2.14). Previously, *Salmonella*-infected HeLa cells have been reported to repress various members of the let-7 family of miRNAs with important implications on the expression of specific pro- and anti-inflammatory cytokines (Schulte et al., 2011). In addition, members of the miR-15 family were shown recently to be suppressed by invading *Salmonella* to trap the host cell in an invasion-favoring cell cycle stage (Maudet et al., 2014). Here, both let-7 and miR-15 family members appeared to be down-regulated in invaded HeLa-S3 cells (not shown). However, despite the possibility to detect (abundant) cellular miRNAs and to quantify their expression changes using Dual RNA-seq, the approach is not tailored for an in-depth miRNA expression profiling. In particular, due to the fragmentation of total RNA samples prior to cDNA synthesis (Fig. 2.1) and the computational removal of short reads (Fig. 2.9) the established protocol counter-selects against short transcripts. Thus, for a more comprehensive miRNA expression screen, library construction following classical protocols dedicated to miRNAs (e.g. size fractionation and enrichment of short RNA species) remains the strategy of choice.

2.9 Concluding remarks

In summary, Dual RNA-seq has been established for a *Salmonella* infection model of different human cell lines. Parameters that have been evaluated for this system include time points after infection to be sampled, different fixation protocols, sequencing depth requirements, read mapping, reproducibility, and the minimal number of cells needed. The pilot Dual RNA-seq experiment demonstrates that the simultaneous detection of all major transcript classes from two fundamentally different transcriptomes can be achieved and used to reliably profile gene expression in both organisms. Organellar RNA such as mitochondrial transcripts and non-coding genes can be also analyzed using this approach. However, as is true for conventional RNA-seq the accuracy depends largely on the abundance (the RPKM) and the coverage of a given transcript.

Findings obtained from the pilot experiment are promising, especially with respect to intracellular *Salmonella* sRNA profiles which unveiled several poorly characterized yet strongly activated candidates. Their induction inside the host cell suggests that these sRNAs could serve important functions for *Salmonella* virulence. The third chapter of this work is therefore dedicated to the functional characterization of one of such intracellularly induced sRNAs, STnc440, and its role in *Salmonella*-host interplay.

3. The STnc440 sRNA regulates *Salmonella* virulence gene expression with impact on the host response

Bacteria possess many regulatory sRNAs, several of which have been suggested to be important for virulence. However, standard infection assays or virulence screens provide little mechanistic insight into how the identified genes contribute to pathogenesis. Here – combining whole-transcriptome studies with classical bacterial genetics and biochemical approaches – the function of the post-invasion induced STnc440 sRNA was characterized, suggesting it to be a novel riboregulator of *Salmonella*'s virulence programs.

3.1 STnc440 inhibition coincides with a fitness defect in different infection models

The pilot Dual RNA-seq experiment identified *Salmonella* STnc440 as the strongest induced sRNA inside HeLa-S3 cells (Fig. 2.24a). Importantly, STnc440 was also activated in human monocyte-like THP-1 cells (second most up-regulated sRNA at both time points; Fig. 2.24b), suggesting its induction is a generic feature of *Salmonella*'s intracellular lifestyle.

STnc440 is a conserved ~80 nt-long *Salmonella*-specific sRNA that is absent from the genome of other Gram-negative bacteria. Within the genus *Salmonella*, however, its primary sequence is highly conserved pointing to a functional sRNA (Fig. 3.1). STnc440 has previously been predicted to be a sRNA in a bioinformatics screen (Pfeiffer et al., 2007) and was validated by Northern blot analysis (Sittka et al., 2008), but has not yet been functionally characterized. Recent data from the lab indicated that STnc440 gets activated through the key regulatory system of intracellular virulence, PhoQ/P, and represses the important invasion factors SopE and SopE2 *in vitro* (Y. Chao, personal communication). However, the role played by this sRNA during *Salmonella*'s interaction with host cells – i.e. the conditions when maximum expression is reached – remained elusive.

In a previous genome-wide mutagenesis screen for *Salmonella* virulence genes (Chaudhuri et al., 2013), disruption of STnc440 expression was found to result in a virulence defect in pig and cattle models of *Salmonella* infection (Fig. 3.2). Several classical proteinaceous virulence factors from the two major pathogenicity islands gave rise to a stronger phenotype when deleted (e.g. SsrB, SopE, SopB). Nevertheless, the effect of STnc440 inhibition was comparable to that of several well-established SPI-2 effectors such as SopD2 (Brumell et al., 2003; Schroeder et al., 2010). As a negative control, disrupting the locus of the conjugal transfer regulator FinO – that

has not been related to infection and virulence – had no effect on *Salmonella* pathogenicity. Interestingly, the *IsrC* sRNA, which has recently been shown to be activated during macrophage infection (Padalon-Brauch et al., 2008), did not compromise *Salmonella* virulence in these models.

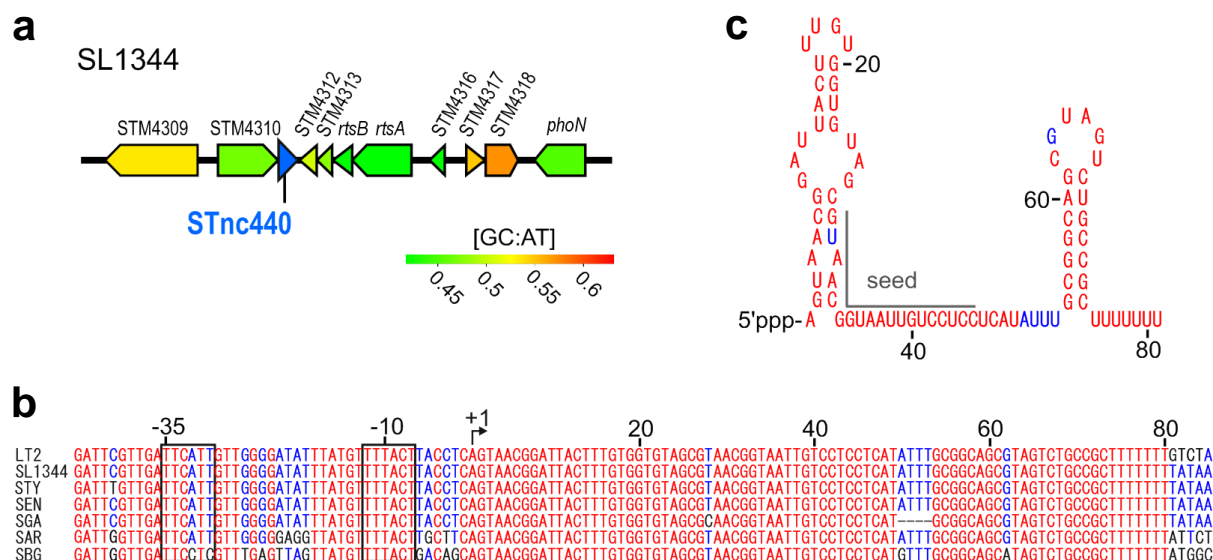


Figure 3.1|The *STnc440* locus. **a**, *STnc440* is a *Salmonella*-specific sRNA on an AT-rich pathogenicity island. **b**, Sequence alignment of the *STnc440* loci of different *Salmonella* species and serovars. “*STY*”: *S. Typhi*, “*SEN*”: *S. enteritidis*, “*SGA*”: *S. gallinarum*, “*SAR*”: *S. arizonae*, “*SBG*”: *S. bongori*. Conserved ribonucleobases are labeled in red, non-conserved bases are written in blue. The numbers indicate the position relative to the 5′ end of *STnc440* (+1 position). **c**, Secondary structure prediction of *STnc440*. Color-coding and numbering as in panel **b**.

In standard cell culture infection experiments, deleting (strain Δ *STnc440*) or complementing *STnc440* (strain Δ *STnc440*+) in the *Salmonella* chromosome had almost no effect on invasiveness (Fig. 3.3a). However, *Salmonella* mutants devoid of *STnc440* expression displayed a slight proliferation defect during the late stage of infection of HeLa-S3 cells (Fig. 3.3b). Milder effects of the inhibition of *STnc440* in cell culture models compared to its *in vivo* phenotype exemplify the limitations of *in vitro* infection systems to identify virulence genes: These simplistic, two-dimensional models that are typically based on a single cell type cannot fully recapitulate the complex environment that a pathogen would normally encounter *in vivo*. In combination with the phenotype of *STnc440* disruption in animal models, however, the findings suggest this sRNA to contribute to *Salmonella* pathogenesis via a yet unknown mechanism.

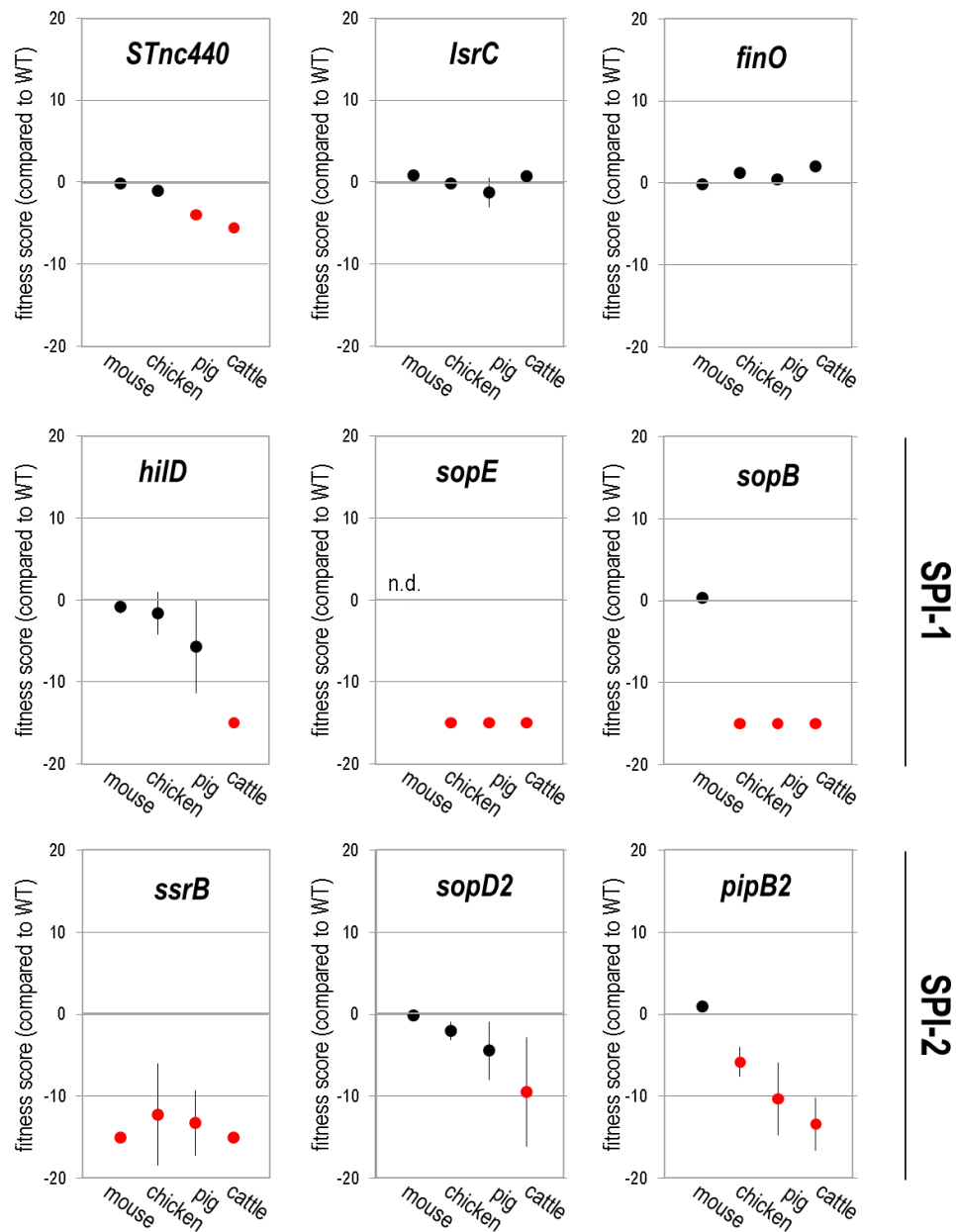


Figure 3.2|The *STnc440* locus contributes to virulence in pig and cattle models of Salmonellosis. Visualization of TraDIS (transposon directed insertion-site sequencing) data for the *STnc440* locus and those of several known virulence genes (data extracted from Chaudhuri et al., 2013). For genes that were hit by multiple transposons, the mean effect and deviation bars are shown. Statistically significant ($p < 0.05$) virulence defects compared to wild-type *Salmonella* are labeled in red. In the screen in mouse, *sopE* was not hit by a transposon insertion (not detected: 'n.d.').

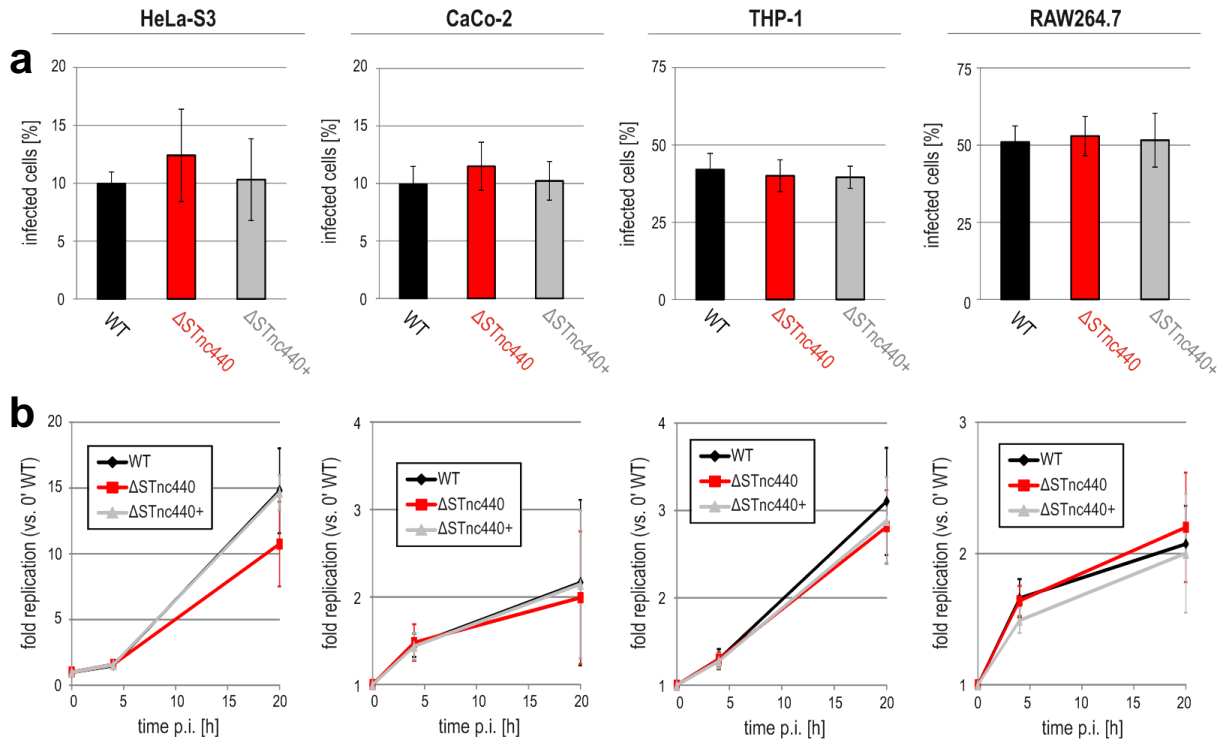


Figure 3.3/Mild effect of STnc440 deletion on intracellular replication within HeLa-S3 cells. a, Standard invasion assays for wild-type, Δ STnc440 and Δ STnc440+ *Salmonella* with human HeLa-S3 (m.o.i. 5), CaCo-2, undifferentiated THP-1 and murine RAW264.7 cells (all m.o.i. 10). Invasion was measured 10 min p.i. by flow cytometry. **b,** Intracellular replication kinetics for the same strains and host cell types. The increase in the GFP intensity of infected cells over time was monitored by flow cytometry and is expressed as fold-change compared to the 0 h time point (as explained in Fig. 2.4). The data represent the mean \pm SD from three independent biological replicates.

3.2 STnc440 expression is rapidly induced after invasion of various host cell types

To comprehensively profile gene expression in intracellular *Salmonella*, a temporally highly resolved Dual RNA-seq experiment was performed (~100-150 M reads per sample; biological triplicates). Samples were taken prior to and 2, 4, 8 and 16 h after infection of HeLa-S3 cells with wild-type *Salmonella* at a m.o.i. of 5 (Fig. 3.4; Tab. 3.1).

Using a cut-off RPKM of ≥ 20 , 106 known (i.e. sRNAs that have previously been detected by Northern blot) and 74 candidate *Salmonella* sRNAs (i.e. sRNAs that have been computationally predicted) were detected (Fig. 3.5). Overall, intracellular sRNA expression levels matched well to the findings obtained from the pilot study (chapter 2). For example, an activation of the FUR-regulated sRNA RyhB (Masse and Gottesman, 2002; Padalon-Brauch et al., 2008) shows again how intracellular *Salmonella* encounter iron limitation and induction of OmrA/B indicates the presence of bacterial surface stress (Beisel and Storz, 2010). In addition, sRNAs known to be controlled by the master regulators of virulence followed the expected kinetics (see Fig. 2.19).

For example, InvR and DapZ sRNAs, which are activated by the major SPI-1 activator HilD (Chao et al., 2012; Pfeiffer et al., 2007), were repressed rapidly after host cell invasion (Fig. 3.5).

Table 3.1 *cDNA libraries generated for the temporally resolved Dual RNA-seq experiment.*

#	library name	<i>Salmonella</i> strain	host cell type	time point	total reads	replicates
1	0 h	WT	HeLa-S3 (as carrier RNA)	0 h	~100 M	3
2	2 h mock GFP-	-	HeLa-S3	2 h	~150 M	3
3	2 h WT GFP+	WT	HeLa-S3	2 h	~150 M	3
4	4 h mock GFP-	-	HeLa-S3	4 h	~150 M	3
5	4 h WT GFP+	WT	HeLa-S3	4 h	~150 M	3
6	8 h mock GFP-	-	HeLa-S3	8 h	~100 M	3
7	8 h WT GFP+	WT	HeLa-S3	8 h	~100 M	3
8	16 h mock GFP-	-	HeLa-S3	16 h	~100 M	3
9	16 h WT GFP+	WT	HeLa-S3	16 h	~100 M	3

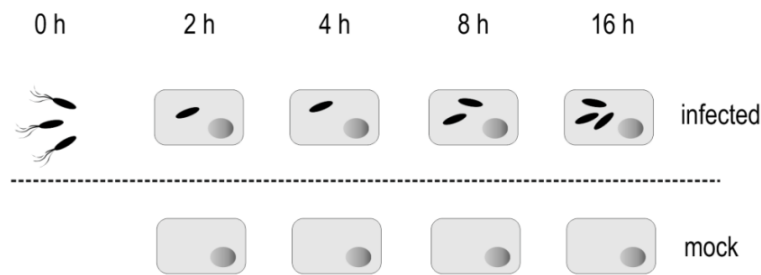


Figure 3.4 *Overview of the high-resolution Dual RNA-seq time-course of HeLa-S3 infection. Four individual time points post-invasion of HeLa-S3 cells were sampled (three replicates). Each time point analyzes invaded host cells only, but not the non-infected bystanders (in contrast to the pilot experiment described in chapter 2). Mock-treated cells were used as reference controls for each time point. As a bacterial reference, extracellular *Salmonella* were included (0 h).*

Importantly, STnc440 remained the most activated sRNA in intracellular *Salmonella*. Its induction kinetics detected by Dual RNA-seq was independently confirmed by qRT-PCR on RNA samples from intracellular *Salmonella* (Fig. 3.6a). Northern blot analysis revealed that full activation of STnc440 as seen for intracellular *Salmonella* can be recapitulated by growing the bacteria in synthetic SPI-2-inducing medium (Lober et al., 2006) *in vitro* (Fig. 3.6b).

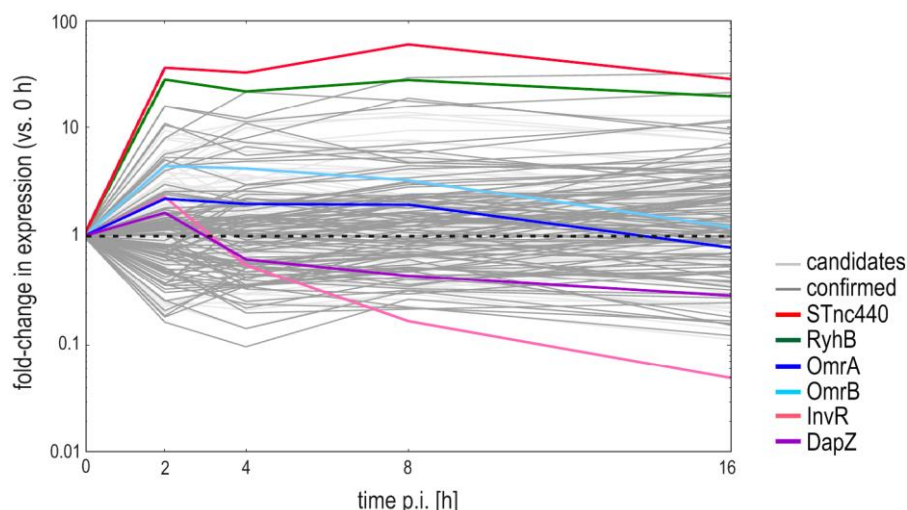


Figure 3.5/Intracellular sRNA expression profile of wild-type *S. Typhimurium* during the time-course of infection of HeLa-S3 cells. All sRNAs detected with a RPKM ≥ 20 at all time points are shown. sRNAs referred to in the main text are labeled in color, the rest are shown as grey lines (dark grey: confirmed sRNAs; light grey: predicted candidates). Data shown comprise three replicates.

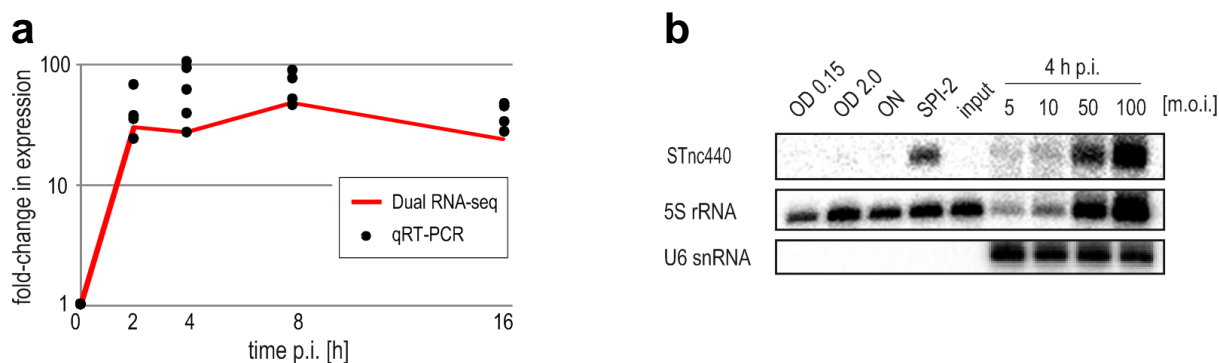


Figure 3.6/STnc440 expression is strongly induced during infection. **a**, STnc440 activation after invasion of HeLa-S3 cells as revealed by Dual RNA-seq (red graph; three replicates; see Fig. 3.5) can be reproduced by qRT-PCR measurements (black dots; each dot represents a single out of four to six biological replicate experiments). qRT-PCR normalization was performed using the constitutively expressed GFP mRNA as a reference. **b**, Northern blot detection of STnc440 under defined in vitro conditions or 4 h after HeLa-S3 infection. Salmonella 5S rRNA and human U6 snRNA serve as loading controls. 'Input' refers to bacteria grown in LB to OD₆₀₀ of 2 and then shifted to host cell medium DMEM.

We aimed to assess the expression levels of STnc440 in different cell types. Besides its induction inside HeLa-S3 cells, Dual RNA-seq at a lower depth (1-5 M reads) indicated STnc440 to be amongst the strongest induced sRNAs in all further tested cell types (Fig. 3.7). These included primary murine bone marrow-derived macrophages (BMDMs), *in vitro*-differentiated THP-1 cells, and human epithelial colorectal adenocarcinoma (CaCo-2) cells. Although the levels of full induction differed between individual cell types, the collective data strongly supports the rapid up-regulation of STnc440 to be a generic feature of the intracellular lifestyle of *Salmonella*.

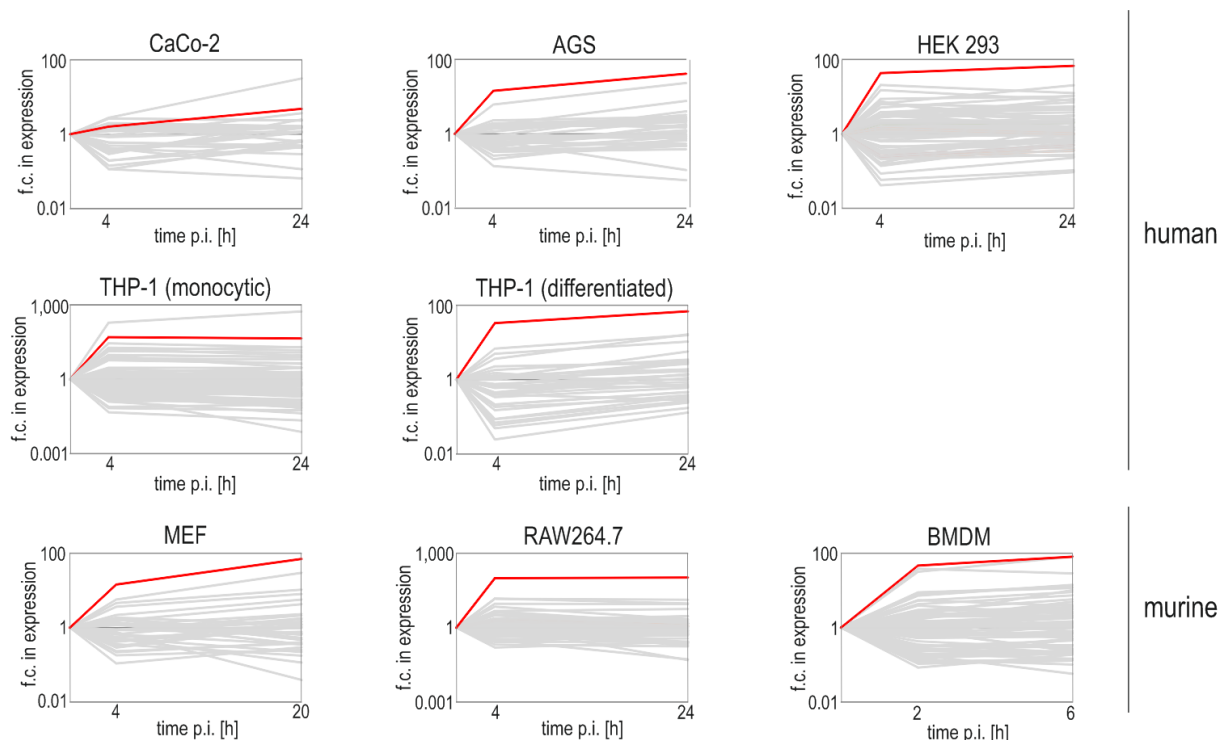


Figure 3.7/STnc440 is strongly induced after the invasion of diverse human and mouse cell types. Cells were infected with wild-type *Salmonella* for the indicated time points, invaded host cells were FACS-sorted and their total RNA was sequenced on Illumina's MiSeq platform to each 1-5 M reads. THP-1 cells either were ('differentiated') or were not ('monocytic') differentiated by phorbol 12-myristate 13-acetate (PMA) treatment for 3 d prior to infection. For all cell types, infection was established at a m.o.i. of 10 and the indicated time points were taken. Expression profiles of all detected sRNAs are plotted (grey) and STnc440 kinetics are labeled red. Fold-changes in expression are relative to extracellular *Salmonella*. In each case, data were derived from a single experiment.

Together, STnc440 represents a paradigm example of a potential, novel virulence-related sRNA: The strong sequence conservation after its horizontal acquisition (Fig. 3.1), its high activation inside various host cell types (Fig. 3.7), and importantly its selection in previous transposon screens for intestinal colonization factors in pig and calf (Chaudhuri et al., 2013) (Fig. 3.2), collectively support a role of STnc440 in bacterial pathogenesis.

3.3 Regulation of *Salmonella* virulence genes by STnc440

Preliminary work in the lab suggested STnc440 to suppress the SPI-1 invasion effectors SopE and SopE2 in a direct manner (Y. Chao, personal communication). However, standard infection screens with diverse cell culture models of *Salmonella* infection did not reveal a defect in host cell invasion in the absence of STnc440 (Fig. 3.3a). In contrast, we detected a mild replication defect of Δ STnc440 mutants at 20 h after HeLa-S3 invasion (Fig. 3.4b). This rather points at an

effect that this sRNA might exert in its intracellular state, e.g. by acting on SPI-2. Indeed, inhibition of STnc440 expression in previous transposon mutagenesis screens (Chaudhuri et al., 2013) mirrored the outcome of the disruption of classical SPI-2 effectors such as SopD2 or PipB2 (Fig. 3.2). Lastly, a preliminary Dual RNA-seq time-course experiment of the infection of HeLa-S3 cells with the Δ STnc440 strain hinted at a potential influence of this sRNA on SPI-2 expression (not shown).

3.3.1 De-repression of SPI-2 genes in the absence of STnc440

For the purpose of profiling global gene expression comparatively between wild-type and Δ STnc440 *Salmonella* with moderate sequencing power, an *in vitro* transition experiment was designed wherein *Salmonella* was shifted from SPI-1-inducing to SPI-2-inducing conditions. This was intended to reconstitute the early stages of host cell invasion and intracellular proliferation *in vitro*. The two *Salmonella* strains were thus grown in LB medium to an OD₆₀₀ of 2, a condition under which SPI-1 is strongly activated (Kroger et al., 2012, 2013). Subsequently the bacteria were pelleted and shifted to SPI-2 medium that mimics the intracellular environment (Lober et al., 2006). Total RNA samples were taken at 0 h (i.e. prior to the shift) or 1 h and 2 h after the medium shift and used to prepare cDNA libraries for RNA-seq.

A functional analysis for enriched pathways compiled from the KEGG database and the literature suggested SPI-2 and the co-regulated SPI-5 (see Fig. 2.20) virulence locus to become gradually hyper-activated over time in *Salmonella* lacking STnc440 (Fig. 3.8). In addition, SPI-1 activation was slightly increased in the sRNA deletion mutant; mostly due to a de-repression of the STnc440 targets SopE/E2 (Y. Chao, personal communication). The pathways “*Salmonella* infection” and “bacterial invasion of epithelial cells” comprise many virulence genes included in the above-mentioned virulence programs and consequently, were also over-activated. Besides these virulence pathways, metabolism-associated GO-terms (such as histidine metabolism) were enriched.

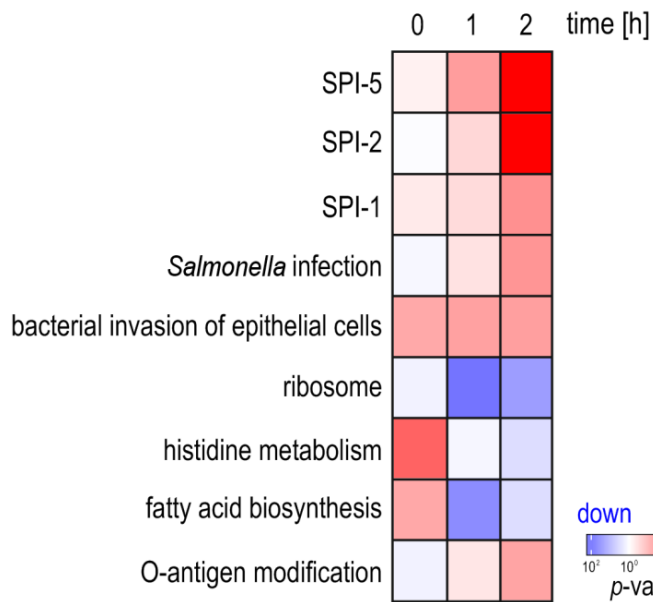


Figure 3.8/Salmonella pathway analysis for the comparative RNA-seq experiment between wild-type and Δ STnc440 mutants *in vitro*. Pathways enriched in differentially expressed genes at any given time point after the medium shift from SPI-1- to SPI-2-inducing conditions are displayed. Red: pathway over-activated in the mutant vs. wild-type; blue: under-activated. Color intensity increases with increased significance (hyper-geometric test). Pathways analyzed were compiled from the KEGG database and the literature. The analysis comprises two replicate experiments.

To test whether the STnc440-dependent SPI-2 repression observed *in vitro* would be detectable also in *Salmonella* growing inside their host, HeLa-S3 cells were infected with wild-type, STnc440 deletion or *trans*-complementation strains. Host-pathogen lysates were prepared prior to or 2 and 4 h after infection, total RNA was isolated and the expression of individual SPI-2 genes was measured by qRT-PCR (Fig. 3.9). The repressive effect of STnc440 on the selected SPI-2 mRNAs, including that of the secreted effector SteC, was subtle but significant also *in vivo*. Importantly, *trans*-complementation of the sRNA gene restored wild-type expression profiles. This supports the notion that STnc440 acts to repress SPI-2 induction during the early phases after invasion. Further qRT-PCR measurements suggested that STnc440 may act upstream of the SPI-2 master transcription factor, SsrB, without affecting its own activator, PhoP, or the master regulator of SPI-1 genes, HilD (Fig. 3.9).

To assess whether the STnc440-mediated influence on SPI-2 expression manifests itself also on protein level, the abundance of the representative SPI-2 effector SteC was monitored *in vitro* at five different time points after the SPI-1 to SPI-2 shift by Western blot analysis (Fig. 3.10). In line with the transcriptomic data, Δ STnc440 bacteria displayed a pre-mature accumulation of this effector, as compared to wild-type and *trans*-complemented *Salmonella*.

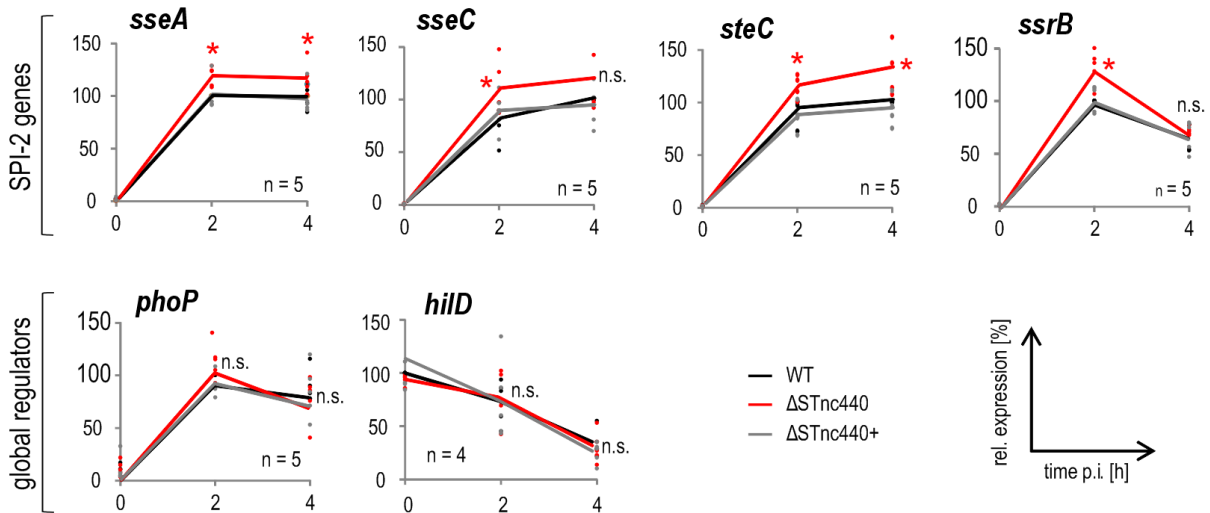


Figure 3.9|Influence of STnc440 on SPI-2 expression in vivo. qRT-PCR measurements during the early stages of HeLa-S3 infection support the de-repression of SPI-2 genes in the Δ STnc440 background (red) as compared to both wild-type (black) and Δ STnc440+ (grey) strains. Normalization was performed using GFP mRNA as a reference. Dots represent individual biological replicate experiments (n) and the solid lines indicate their mean. A Mann-Whitney U test was performed to test for significance (*: $p < 0.05$; n.s.: $p > 0.05$ between wild-type and Δ STnc440 strains).

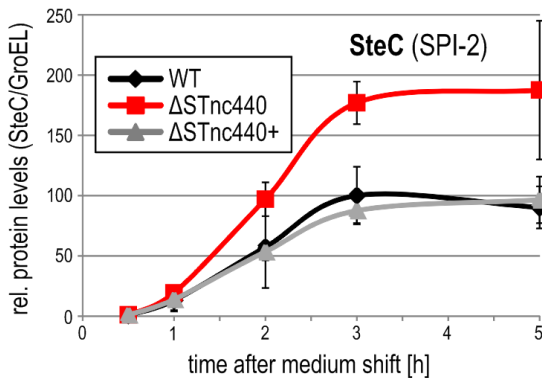


Figure 3.10|STnc440 affects the accumulation of the SPI-2 effector protein SteC in vitro. Western blot analysis of a tagged *steC::flag* *Salmonella* strain shifted from SPI-1- to SPI-2-inducing conditions. SteC levels were normalized to GroEL. Graphs and error bars represent the mean \pm SD from three independent time-course experiments.

3.3.2 Pulse-expression experiments identify STnc440 targets

Since STnc440 associates with Hfq (Chao et al., 2012; Sittka et al., 2008), it is likely to regulate mRNAs by base pairing (Vogel and Luisi, 2011). Thus, in order to identify direct targets of this sRNA, a previously described pulse-expression approach was applied (Masse et al., 2005; Papenfort et al., 2006). Briefly, the sRNA of interest is transiently over-expressed for a short time period (5-10 min) from an arabinose-inducible promoter and the global abundance of mRNAs is monitored in the presence or absence of the sRNA. The expression of direct mRNA targets is expected to change while indirect effects are minimized due to the short time frame of induction.

To identify STnc440 targets in conditions mimicking the host cell milieu (i.e. conditions when STnc440 is maximally expressed), the sRNA was transiently over-expressed in *Salmonella* growing in synthetic SPI-2-inducing medium (Lober et al., 2006). However, to take into account that an artificial medium might not perfectly reflect the conditions found intracellularly (Ortega et al., 2012), an experiment was also designed to transiently induce STnc440 expression in bacteria within their host cell niche. Previous work suggested that supplementing the host cell medium with L-arabinose would be sufficient for the inducer to traverse the host's plasma membrane and diffuse into the cytosol (Loessner et al., 2007). We thus added L-arabinose into the supernatant of HeLa-S3 cultures which had been infected for 4 h with GFP-expressing *Salmonella* containing the STnc440 over-expression plasmid. At 2, 5, 10 and 20 min post-induction, total RNA samples were taken and STnc440 expression was monitored by qRT-PCR (Fig. 3.11a). This revealed that in wild-type or Δ STnc440 *Salmonella* transformed with the empty control plasmid, STnc440 levels were as expected: A ~100 fold induction was measured in the wild-type background at 4 h p.i. compared to the levels in extracellular bacteria prior to infection (Fig. 3.11a, black graph), in line with previous findings (Fig. 2.24, Fig. 3.6a). Conversely, the sRNA was not detected in the Δ STnc440 background (Fig. 3.11a, grey graph). In a Δ STnc440 strain transformed with the STnc440-containing expression plasmid, however, the sRNA was lowly expressed in the absence of the inducer (0 min), but rapidly up-regulated to up to ~500 fold as soon as 5 min post-induction (Fig. 3.11a, red curve).

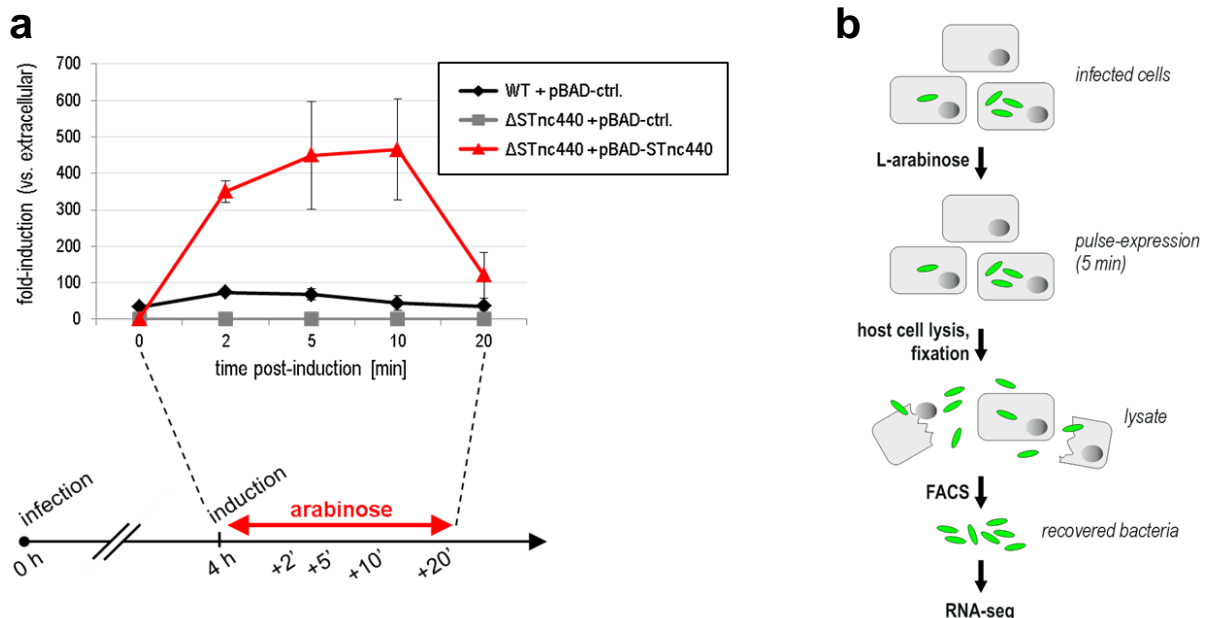


Figure 3.11|In vivo sRNA pulse-expression establishment. **a**, 4 h after the infection of HeLa-S3 cells, STnc440 over-expression was induced from the pBAD plasmid by the addition of L-arabinose into the host cell medium. Up-regulation of STnc440 (red) was monitored by qRT-PCR and compared to its expression in the wild-type (black) or Δ STnc440 strain (red) containing the empty pBAD vector. **b**, Work-flow for the enrichment of bacterial RNA after in vivo pulse-expression.

Therefore, Δ STnc440 strains either harboring the empty control or the STnc440 expression plasmid were used to infect HeLa-S3 cells for 4 h and subsequently L-arabinose was added into the medium for 5 min (Fig. 3.11b). Infected host cells were harvested and lysed by detergent treatment on ice. The released bacteria were pelleted by centrifugation. To freeze their transcriptome, *Salmonella* cells were fixed in RNAlater (see section 2.2). In order to further reduce the host RNA background, fixed *Salmonella* were enriched via FACS by gating for the GFP-positive population of released bacteria. RNA was isolated from the collected samples and analyzed by RNA-seq. The *in vivo* data was used to complement the results from pulse-expression in SPI-2 medium (Fig. 3.12). The combined datasets predicted the mRNAs of the enzyme glutaredoxin 1 (GrxA) and the transcriptional regulator cyclic AMP receptor protein (CRP) as high-confidence repressed targets of STnc440. Furthermore, the mRNA of the bacteriolytic toxin entericidin B (EcnB) was included in the downstream analyses, as strong repression was detected in SPI-2 medium (Fig. 3.12). Note that the mRNAs for the two SPI-1 effectors SopE/E2 – which are direct STnc440 targets under SPI-1 conditions (Y. Chao, personal communication) – were lowly abundant under the *in vivo*-like conditions sampled and consequently not identified here as targets.

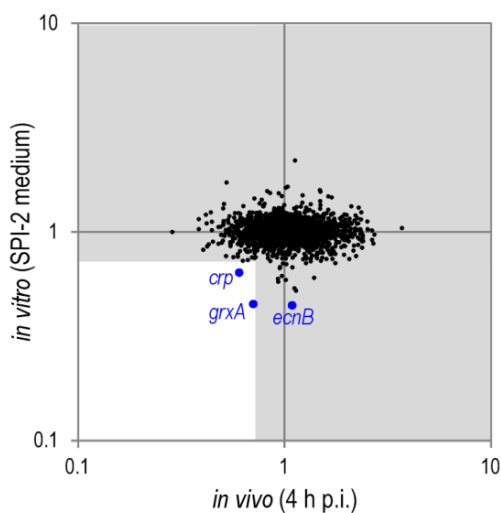


Figure 3.12/Pulse-expression experiments to identify target genes of STnc440. The sRNA was over-expressed under SPI-2-inducing conditions *in vitro* or 4 h after HeLa-S3 infection. In each case, total RNA samples were taken 5 min after induction and analyzed by RNA-seq (two independent replicates per condition). Axes represent the fold-changes in mRNA abundance between strains harboring the empty control plasmid and the STnc440-containing plasmid. The mRNAs that were confirmed to be directly targeted by STnc440 (see Fig. 3.13 and 3.14) are labeled blue.

Western blot analysis was used to assess whether STnc440 expression affected the protein levels of these three putative targets (Fig. 3.13). Under SPI-2 conditions, a time-course of protein abundance revealed a rapid clearance of GrxA, CRP and EcnB following the expression of STnc440 from a plasmid under its native promoter. The level of the unrelated GroEL protein remained unaffected.

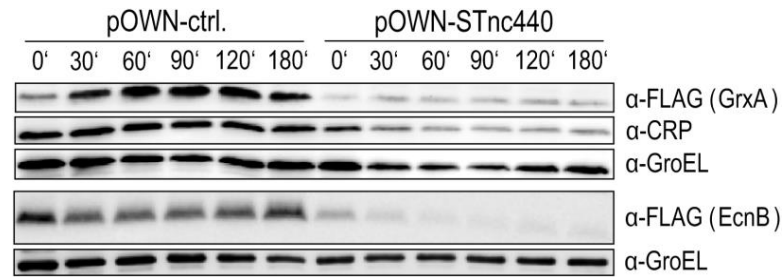


Figure 3.13|Western blot analysis for STnc440 targets. Protein levels upon the expression of STnc440 from a plasmid containing the sRNA under its native promoter (pOWN-STnc440) in strains with FLAG-tagged *grxA* or *ecnB* that were shifted from LB to SPI-2 medium. At defined time points after the shift total protein samples were taken and analyzed by Western blot for the high-confidence STnc440 targets identified in Fig. 3.12. GroEL served as a loading control.

In silico approaches were used to search for potential target sites within the *grxA*, *crp* and *ecnB* mRNAs (Rehmsmeier et al., 2004). In the case of *grxA* a high confidence interaction site between the region around the start codon of the mRNA and STnc440 could be predicted (Fig. 3.14a). As validated using a translational GFP reporter construct, targeting of *grxA* by STnc440 is direct (Fig. 3.14b). Intriguingly, regulation was diminished for a point mutant of STnc440, in which two bases within the seed region (see Fig. 3.1) were exchanged (variant STnc440*). Probably due to the extensive RNA-RNA interactions predicted to form between the two transcripts, exchanging a single ribonucleobase within the seed (STnc440**) was insufficient to abolish regulation.

Two *crp* mRNA isoforms have been reported (Kroger et al., 2012) that differ with respect to their 5' UTR: The primary TSS gives rise to a mRNA with a shorter 5' UTR, while an isoform with a 5' UTR extended by 76 nt is produced from the secondary TSS (Fig. 3.14c). For both *crp* isoforms suitable GFP reporter constructs were generated and the effect of STnc440 over-expression on them was analyzed (Fig. 3.14d). This indicated that primarily the long isoform is regulated by STnc440 (but not STnc440*), even though slight repression was also observed for the shorter transcript. This suggests that the primary target site is located in the region exclusive to the long *crp* mRNA variant. Indeed a weak but possible interaction site with the STnc440 seed region could be identified within the respective 76 nt window (Fig. 3.14c). Additionally, there might be (a) further targeting site(s) shared by both *crp* isoforms. For *ecnB* mRNA several but weak interaction sites were predicted, however, these were not further addressed here.

Western blot analysis supported that efficient clearance of GrxA and CRP proteins depends on an intact seed region of STnc440 (Fig. 3.14e). The ectopic expression of the wild-type sRNA in bacteria grown under SPI-2 conditions decreased GrxA and CRP levels, whereas over-expression of the point-mutated version of STnc440 (variant STnc440*) had little effect on target protein levels.

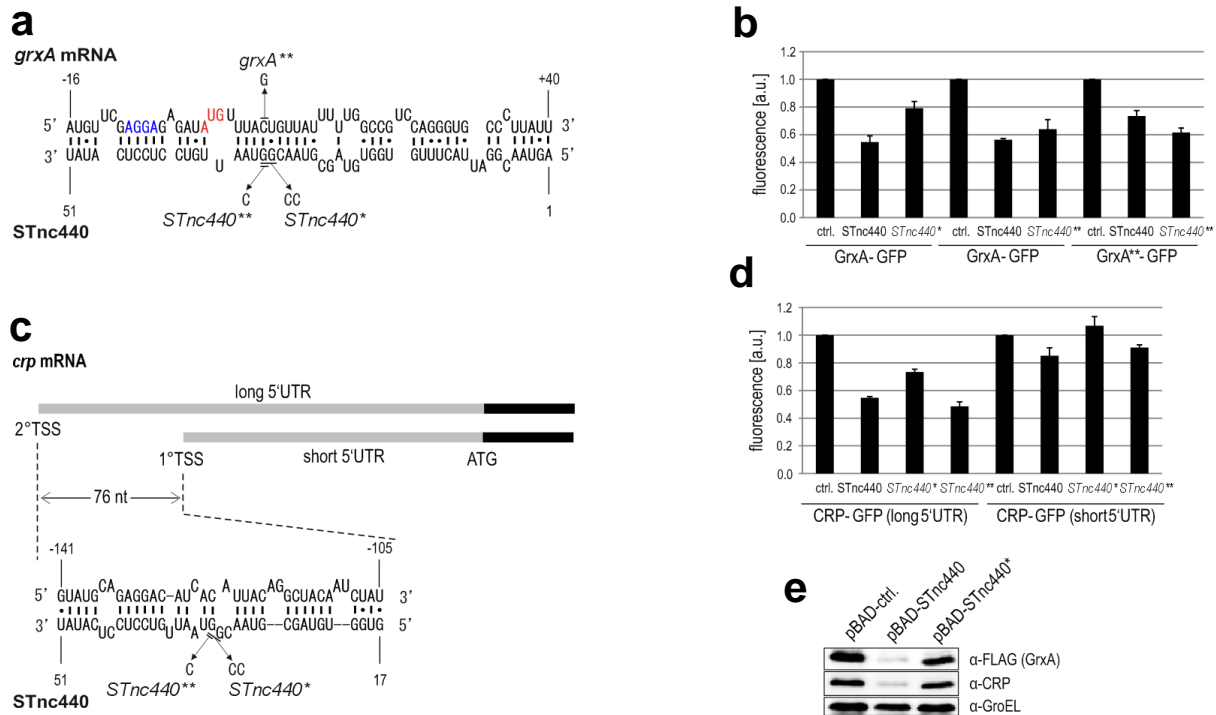


Figure 3.14|Direct targeting of *grxA* and *crp* mRNAs by STnc440. **a**, Predicted RNA duplex formation between STnc440 and the 5' part of *grxA*. Point mutants analyzed in panel **b** are indicated. The ribosome binding site and start codon of *grxA* are marked in blue and red, respectively. **b**, Validation of direct targeting of the *grxA* mRNA by the seed region of STnc440 using translational *grxA::gfp* reporter gene fusions. For these reporters, the region -34 to +90 of *grxA* (relative to the position of the start codon) was fused to the coding sequence of *gfp*. *Salmonella* strains containing both a GFP reporter plasmid and a sRNA over-expression vector were grown overnight in LB and analyzed by flow cytometry. The error bars indicate standard deviation from the mean based on triplicate measurements. **c**, Putative RNA duplex formation between STnc440 and *crp*. Two isoforms of *crp* mRNA exist that differ with respect to their 5' UTR. Point mutations introduced in the STnc440 seed (panel **d**) are indicated. **d**, Validation of direct targeting of the long *crp* mRNA isoform by the seed region of STnc440 using suitable translational *crp::gfp* reporter fusions (as in panel **b**). **e**, Western blot analysis for GrxA and CRP proteins upon the over-expression of either wild-type STnc440 or the point-mutated sRNA variant (STnc440*).

In summary, STnc440 seems to have widespread regulatory effects that are mediated in either a direct or indirect fashion: Under SPI-2 conditions (*in vivo* and *in vitro*-reconstituted), STnc440 targets and represses *grxA* and *crp* mRNAs. Also the bacteriolytic toxin EcnB appears to be amongst STnc440's targets although no validation experiments were performed for this factor. In addition, STnc440 seems to exert a global repressive effect on the SPI-2 locus, likely via an indirect mechanism. Since GrxA and CRP affect intracellular replication by activating *Salmonella* virulence genes (Bjur et al., 2006; Chen et al., 2010; Yoon et al., 2009), they are considered putative factors through which STnc440 might signal to SPI-2. Likewise, EcnB levels have recently been shown to be regulated between proliferating and resting *Salmonella in vivo* (Claudi et al., 2014).

Thus, to address if GrxA, CRP or EcnB were involved in the effect of STnc440 on SPI-2 activity, appropriate *Salmonella* strains were subjected to the *in vitro* SPI-1 to SPI-2 transition assay described above. As a proxy for SPI-2 activity, induction of a transcriptional reporter fusion comprising the SsrB-activated SPI-2 promoter of *ssaG* fused to the coding sequence of *gfp* was monitored (Fig. 3.15). Again, the medium-switch assay recapitulated the premature activation of SPI-2 transcription in the absence of STnc440 (red curve in Fig. 3.15). Reciprocally, the presence of STnc440 on a multi-copy plasmid impaired SPI-2 gene activation (grey curve in Fig. 3.15). Importantly, genomic deletion of *crp* (but not *grxA*) abrogated and *ecnB* deletion weakened the STnc440 effect on the *ssaG::gfp* fusion, suggesting that STnc440 might signal to SPI-2 through CRP, possibly involving EcnB.

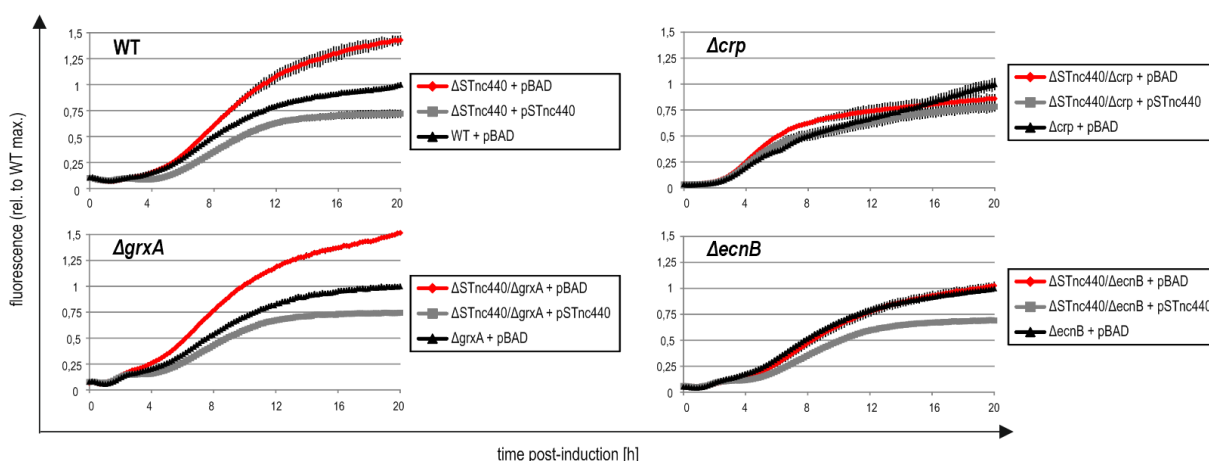


Figure 3.15/STnc440 counter-acts SPI-2 activation through CRP *in vitro*. *Salmonella* strains with endogenous STnc440 expression (black), strains over-expressing (grey) or lacking STnc440 (red) in either the wild-type or a defined deletion mutant background were shifted from SPI-1 to SPI-2 conditions. A transcriptional reporter fusion of the SPI-2 gene *ssaG* to *gfp* was used as a proxy for SPI-2 induction. Time point 0 h refers to the time of the medium shift and the induction of ectopic sRNA expression. Fold-changes are expressed relative to the maximum activation levels of *ssaG* reached in the presence of wild-type STnc440 expression (black curves). The error bars indicate SD from the mean from triplicate measurements.

3.4 JAK/STAT and cytokine signaling are STnc440-affected host pathways

Secreted bacterial effector proteins are directly exposed to the surveillance machinery of the host. Therefore, observations that STnc440 affects the level of *Salmonella* virulence effectors from SPI-1 (Y. Chao, personal communication) and SPI-2 (see above), implied that also the host response to infection could be influenced by the action of this sRNA (Bruno et al., 2009; Hannemann et al., 2013; Uchiya and Nikai, 2005, 2008).

For example, the JAK/STAT (Janus kinase/Signal transducer and activator of transcription) signaling cascade is well known to be a major host cellular target process of many intracellular pathogens, including *Salmonella* (reviewed in Carey et al., 2012). Indeed, preliminary data from a Dual RNA-seq time-course with Δ STnc440 *Salmonella* suggested that this pathway might be differentially affected in the absence of STnc440 (not shown). Within the JAK/STAT pathway members of the Suppressor of cytokine signaling (SOCS) family play a key role as mediators of downstream signaling events. Their expression is typically controlled transcriptionally and thus changes are reflected already on the mRNA level (Yoshimura et al., 2007). Intriguingly, previous studies reported differential *SOCS* expression in a SPI-2 effector-dependent manner (Uchiya and Nikai, 2005, 2008). Here, to assess whether the JAK/STAT pathway was affected by STnc440, expression of key members of the SOCS family was profiled by qRT-PCR (Fig. 3.16). *SOCS* activation was almost exclusively observed in the invaded cells but not the non-infected bystanders (solid vs. dashed lines in Fig. 3.16). Changes in the expression kinetics between wild-type- or mutant-infected cells were observed for all the addressed factors. Most prominently, however, *SOCS3* mRNA was prematurely induced upon the invasion with Δ STnc440 as compared to wild-type bacteria.

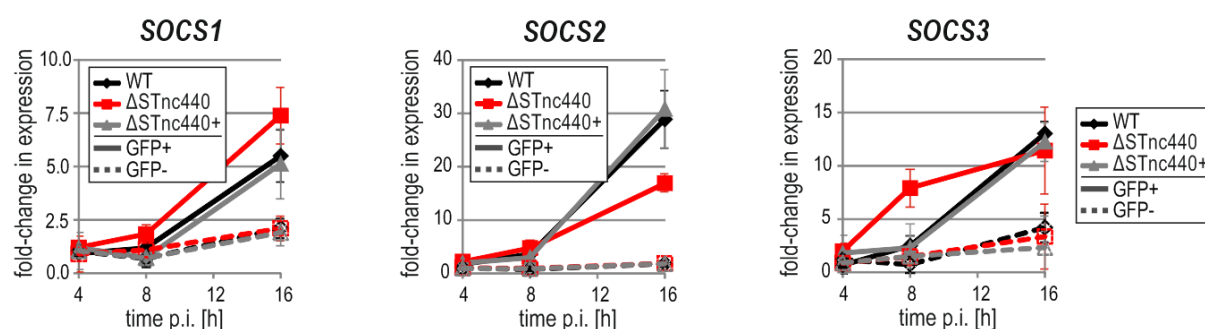


Figure 3.16|Infection with Δ STnc440 *Salmonella* impacts on *SOCS3* induction of the host. qRT-PCR measurements on total RNA samples isolated from FACS-sorted HeLa-S3 cells at the indicated time points after wild-type, Δ STnc440 or Δ STnc440+ infection. U6 snRNA was used for normalization. The data represent the mean \pm SD from three independent biological replicates.

SOCS3 has previously been shown to inhibit tyrosine phosphorylation of the transcription factor STAT3, thereby preventing its activation and nuclear import (Li et al., 2012a). In agreement with this, reduced STAT3 tyrosine phosphorylation in cells infected with Δ STnc440 *Salmonella* compared to wild-type infection was observed by means of Western blot analysis using a Tyr⁷⁰⁵-phospho-specific antibody (Fig. 3.17). As a control, a second, non-canonical phosphorylation site at residue Ser⁷²⁷ (Gough et al., 2009) was unaffected by STnc440.

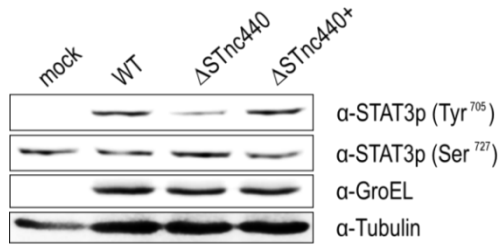


Figure 3.17|STnc440 affects host STAT3 phosphorylation. The phosphorylation status of STAT3 was assessed by Western blot analysis of unsorted HeLa-S3 cells at 16 h p.i. using *Tyr*⁷⁰⁵- or *Ser*⁷²⁷-phospho-specific antibodies against STAT3. Salmonella GroEL served as a bacterial and tubulin as a human control.

A classical readout in profiling inflammatory responses of stimulated cells is to measure cytokine/chemokine levels. As revealed in the pilot Dual RNA-seq experiment (chapter 2) infected HeLa-S3 cells strongly up-regulated the mRNAs for the pro-inflammatory chemokine IL-8 and the IL-10-like, anti-inflammatory cytokine IL-24 (a.k.a. MDA-7) (e.g. Fig. 2.27). Thus, to address whether the STnc440 genotype of invading *Salmonella* would affect the inflammatory response of its corresponding host, qRT-PCR measurements of *IL8* and *IL24* mRNAs at 16 h after infection (i.e. at the maximal height of their expression) were performed (Fig. 3.18a). These data indicate that HeLa-S3 cells infected with mutant *Salmonella* lacking STnc440 hyper-expressed *IL8* while *IL24* induction was reduced as compared to wild-type infection. Of note, these alterations in gene expression were not caused by a differential bacterial load of the individual strains at 16 h p.i. as indicated by a similar *GFP* mRNA to U6 snRNA ratio for all conditions (Fig. 3.18b). Enzyme-linked immunosorbent assay (ELISA) at 20 h p.i. demonstrated increased *IL8* mRNA expression after Δ STnc440 infection resulted in elevated levels of IL-8 protein in the culture supernatant (Fig. 3.18c). All reported alterations reverted to levels seen in wild-type-infected cells upon *trans*-complementation of the sRNA (Fig. 3.16, 3.17 and 3.18).

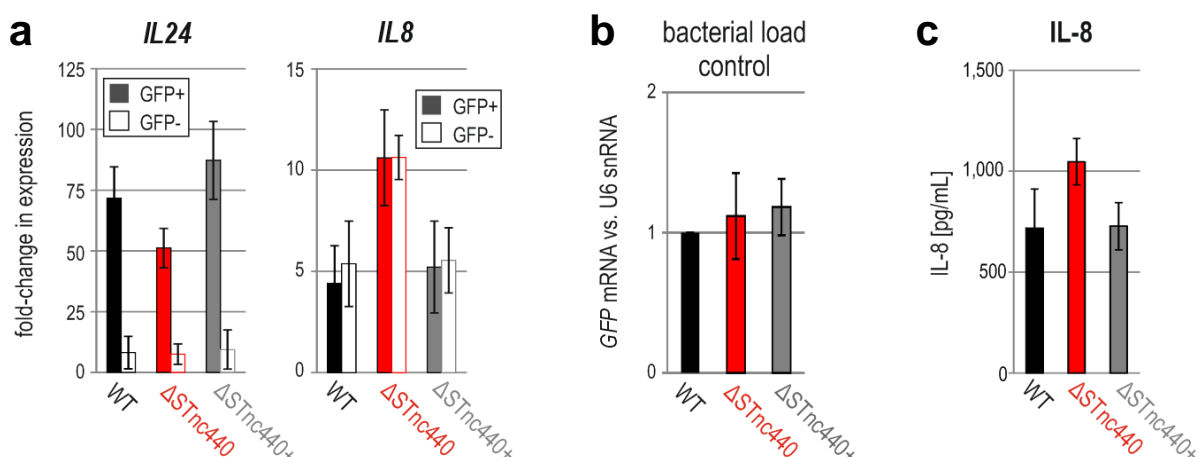


Figure 3.18|STnc440-dependent changes in host cytokine/chemokine levels. **a**, qRT-PCR on RNA from FACS-sorted HeLa-S3 cells at 16 h p.i. with the indicated strains. U6 snRNA was used for normalization. **b**, Differential gene expression events in panel a were not caused by bacterial load variations as judged from the qRT-PCR detection ratio of bacterial *GFP* mRNA to human U6. **c**, ELISA detected elevated concentrations of secreted IL-8 protein in the culture supernatant at 20 h after Δ STnc440 infection. All data represent the mean \pm SD from at least three independent biological replicates.

3.5 Concluding remarks

Based on the results from the pilot Dual RNA-seq screen, STnc440, a strongly induced but poorly studied *Salmonella* sRNA was selected for in-depth characterization. A preliminary Dual RNA-seq screen with the Δ STnc440 mutant was used to pinpoint putative *Salmonella* and host processes affected by this sRNA. In combination with the observed growth defect of the sRNA mutant, this led us to globally profile gene expression during the onset of intracellular proliferation. Under the defined conditions of an *in vitro*-reconstituted growth experiment that mimicked the intracellular milieu, pathway analysis revealed SPI-2 genes to be de-repressed in the absence of STnc440. In parallel, based on the body of previous work on *Salmonella* effectors we monitored whether STnc440 expression would influence important branches of the immune response in the host. The combined data illustrate how STnc440 may orchestrate a complex regulatory network of *Salmonella* virulence gene expression, likely to shape the transition from an invasive to an intracellularly replicating lifestyle. The function of STnc440 in pathogen-host interplay further manifests itself in changes in important signaling cascades of the host. Differences in the activity of the here-assessed (and possibly further) immune pathways between wild-type- and Δ STnc440-infected cells suggest STnc440 might act to indirectly modulate the inflammatory host response.

4. Discussion

4.1 Dual RNA-seq: potential and scope for future improvements

Infection is one of the most intricate interactions between two organisms from different species – especially in a scenario where the pathogen invades its host. Thus, rather than laboriously separating the interaction partners post infection, methods are required that enable the simultaneous analysis of both organisms. In the case of proteins, *in silico* separation of mixed bacterial/eukaryotic peptide samples is routinely performed (reviewed in Rodland et al., 2008 for the example of *Salmonella* infection models). Even dynamic analyses are feasible using stable isotope labeling of amino acids. These approaches are capable of revealing how intracellular pathogens reprogram their proteome from inside the host cell (Rechavi et al., 2010) or identifying secreted bacterial effectors and their cognate host targets (Auweter et al., 2011). However, as our perception of gene expression shifts from a protein-centric to a more RNA-centric view both in pro- (Papenfert and Vogel, 2010) and eukaryotes (Dunham et al., 2012), such parallel techniques need to be transferred to nucleic acids.

On the RNA level, joint analyses have proven more challenging, mostly due to the dramatic differences concerning host and pathogen transcriptomes (Westermann et al., 2012). Recently several studies have reported the application of joint RNA-seq to infection models (Kawahara et al., 2012; Petre et al., 2012; Tierney et al., 2012; Yazawa et al., 2013), however, due to the fundamental differences between bacterial and eukaryotic transcriptomes such analyses have been restricted to hosts infected with eukaryotic parasites. The ubiquitous polyadenylation of eukaryotic mRNAs and the sequence similarity of ribosomal transcripts facilitates the experimental protocol as it enables the poly(A) enrichment or ribosomal depletion for host and pathogen in a single step. Recently a few publications have reported the application of RNA-seq to a model of bacterial infection of plant leaves (Asai et al., 2014; Socquet-Juglard et al., 2013), porcine enterocytes (Vannucci et al., 2013), or human epithelial carcinoma cells (Humphrys et al., 2013). However, these suffered from the lack of an enrichment of invaded cells, resulting in a low bacterial transcriptome coverage, and/or from a shallow sequencing depth, hampering an in-depth profiling of the host response.

Here, a generic protocol for Dual RNA-seq was established that should be transferable to any infection system, eventually even to tissue or animal models. Samples were fixed directly after harvesting; this means that the transcriptome composition will be maintained during all downstream processes such as the enrichment of infected cells. Note, however, that the here described *RNAlater*-based fixation method was optimized for the analysis of robust,

immortalized cell lines. For other sample types – especially primary cells or tissue samples – this treatment might turn out to be too harsh. Thus, further fixatives and protocols should be evaluated in the future. The current protocol for Dual-RNAseq requires as little as 10,000 infected cells. However, as the sensitivity of library construction continues to increase, the time required for cell sorting will decrease and fixation might then be omitted. This strategy is typically used for single-cell RNA-seq studies (Saliba et al., 2014).

Using Dual RNA-seq, sample preparation is simpler than for previous microarray-based approaches as the interacting organisms can be analyzed together and mixed RNA samples are directly converted into cDNA in the absence of further enrichment or depletion steps. As a result, little starting material is required which accelerates sampling and allows for the design of larger-scale experiments (e.g. different host cell types or pathogen strains could be handled in parallel). The many advantages over probe- or tag-based techniques (see Introduction) have rendered conventional RNA-seq the gold standard for single-organism transcriptomics (Wang et al., 2009). As the same advantages are inherent to the parallel analysis of host and pathogen transcriptomes by Dual RNA-seq, it might eventually become an important technique for microbial infection biology.

Here, Dual RNA-seq was performed on Illumina's HiSeq 2000 platform. Assigning up to one lane of a flow cell (~150 M reads) to a single library was found to be reasonable for a Dual RNA-seq experiment that starts from total RNA. The resulting ~0.5-5 M bacterial non-ribosomal reads (depending on the time p.i.) and the ~50 M human non-ribosomal reads appear sufficient for the characterization of both pathogen and host transcriptomes. Differential gene expression studies were possible on both sides and the calculated expression changes were in good accordance with results obtained from independent qRT-PCR or Northern blot experiments. Note, however, that it was initially intended to complement the temporally resolved time-course of *Salmonella* wild-type infection with that of the Δ STnc440 mutant strain. The high degree of reproducibility of strong expression changes between infectious and non-infectious conditions notwithstanding, the subtle changes between wild-type and Δ STnc440 infection were not consistently detected in the sequencing datasets from multiple biological replicates. As a result, the differences did not withstand rigorous significance assessment using state-of-the-art analysis tools. One explanation could be that the HeLa-S3 cell culture model used here might impede such a comparative analysis. That is, we initially selected HeLa-S3 cells as a robust *in vitro* model of infection for the purpose of establishing the Dual RNA-seq technique. However, as a consequence of its simplicity, subtle but potentially meaningful changes might be mitigated or even entirely masked in this model system. In addition, the high degree of variability that accompanies the infection process would probably require more replicates to accurately quantify slight expression changes. Lastly, with the current sequencing depth the Dual RNA-seq method might not be sufficiently sensitive

to quantitatively assess the subtle differences in gene expression between wild-type bacteria and a single sRNA mutant or their respective hosts. Indeed, for a more comprehensive characterization of mammalian transcriptomes ~700 M non-rRNA reads (Blencowe et al., 2009) and for saturated bacterial transcriptomics >10 M non-rRNA reads (Haas et al., 2013) were recently suggested. As an alternative to the sequencing of total cDNA, rRNA may be depleted from the samples prior to library construction in the future. That is, recently a ribosomal depletion kit was launched that enables the joint removal of bacterial and eukaryotic rRNA in a single step (Ribo-Zero Magnetic Gold kit from epicenter). Such an approach promises to reduce sequencing depth requirements for future Dual RNA-seq studies.

4.2 Rewiring the regulatory network of virulence gene expression in *Salmonella*

Making use of ENCODE-compatible HeLa-S3 or THP-1 cells challenged with *S. Typhimurium* as a robust cell culture model of infection, has enabled us to provide deeper insight into the expression programs triggered in pathogen and host during the course of infection. For example, inspecting the bacterial data subset informed about the various types of stresses that the pathogen experiences inside its different host cell niches. Also information about the expression of virulence genes from different programs could be obtained. Intriguingly a previously poorly characterized *Salmonella* sRNA was found to be highly induced inside all host cell types tested.

Using global transcriptomics complemented with standard genetics and biochemical techniques, evidence was collected that this *trans*-acting sRNA operates at the transition of SPI-1 to SPI-2 expression and thus, at the heart of *Salmonella* virulence (Fig. 4.1). In particular, STnc440 is encoded within a pathogenicity island and genomically adjacent to the *rtsA* gene, which encodes a known regulator of SPI-1 genes (Ellermeier and Slauch, 2003). RtsA, however, was dismissed as a transcriptional activator of STnc440 due to the opposing expression patterns of this sRNA and SPI-1 invasion genes. Rather STnc440 may be activated through the major regulatory system of intracellular virulence gene expression, PhoQ/PhoP (Y. Chao, personal communication). In turn, STnc440 regulates specific effectors from SPI-1 by direct base-pair interaction with the respective mRNAs (Y. Chao, personal communication) as well as expression of the bulk of SPI-2 genes via an indirect mechanism, possibly involving CRP.

The STnc440 deletion mutant displayed a replication defect during the late stages of HeLa-S3 infection and genome-wide transposon mutagenesis screens have previously identified insertion mutants with impaired STnc440 expression to have a fitness defect in diverse animal models of Salmonellosis. To some extent, this may be linked to the mis-regulation of SPI-2 expression in the absence of STnc440. It has been reported before that premature SPI-2 induction has an

adverse effect on *Salmonella* growth within its host cell niche and therefore negative regulators of SPI-2, such as YdgT (Coombes et al., 2005) or PmrA (Choi and Groisman, 2013), are required for the contextual activation of virulence factors from this island (Fig. 1.2). The present data suggest that STnc440 contributes to the modulation of SPI-2 activity.

CRP was identified as a possible mediator of this regulation. This factor has previously been suggested to impact *Salmonella* virulence by affecting both SPI-1 (Chen et al., 2010) and SPI-2 (Yoon et al., 2009), via unknown mechanisms. Whereas the presented data also cannot provide a mechanistic link between CRP expression and SPI activity, it suggests that the CRP hub in the regulatory network would provide a temporal gap between PhoP-dependent SPI-2 induction and its indirect, STnc440-mediated repression (Fig. 4.1). This way, virulence effector expression inside the host would be restricted to within a narrow time window – sufficient for optimal host manipulation but avoiding hyper-production of bacterial effectors, thereby potentially minimizing recognition by host cell sensors. Future efforts – using cell culture but also animal models of *Salmonella* infection – should aim to evaluate the here proposed role of STnc440 in *Salmonella* virulence.

The transcriptional regulator CRP senses intracellular glucose levels by means of cyclic AMP binding and controls the expression of several hundred of target genes in the Enterobacteriaceae (Xu and Su, 2009; Zheng et al., 2004). Besides its emerging role in virulence (see above), CRP is involved in carbon catabolite repression, flagellum synthesis and toxin production (Botsford and Harman, 1992). Indeed, pathway analysis during the *in vitro*-reconstituted SPI-1 to SPI-2 switch revealed potential footprints of STnc440-mediated *crp* repression beyond its effect on virulence genes (e.g. histidine metabolism was affected). As the experimental conditions, however, were selected specifically for the purpose of monitoring virulence gene expression it appears likely that STnc440's effect on CRP may be more prominently reflected in the transcriptome of *Salmonella* grown under different conditions. While the exact role of this sRNA for *Salmonella*'s metabolism remains to be elucidated in the future, the obtained data suggest STnc440 could be an example of the emerging theme of sRNAs acting at the interface of bacterial metabolism and virulence (Papenfort and Vogel, 2014).

The results from ectopically expressing STnc440 under *in vivo* conditions hint at an extensive targetome for this sRNA. Besides CRP, glutaredoxin and entericidin were found to be additional high-confidence STnc440 targets. *Salmonella* and *E. coli* have two major redox systems to repair oxidized sulfhydryl groups in cytosolic proteins, namely the thioredoxin and glutathione/glutaredoxin systems (Stewart et al., 1998). Previous studies have uncovered a link between the expression of thioredoxin 1 (TrxA) and *Salmonella* virulence, as a *trxA* deletion strain displayed a virulence defect in BALB/c mice and a cell culture model of murine macrophage-like cells (Bjur

et al., 2006; Negrea et al., 2009). In the same work deletion of *grxA* was found to increase fitness of *Salmonella* in the macrophage model (Bjur et al., 2006).

Recently slowly growing, persistent *Salmonella* within host tissues have been reported to contain elevated levels of EcnB protein compared to actively replicating bacterial cells (Claudi et al., 2014). The fact that the bacteriolytic toxin EcnB is part of a toxin/anti-toxin (TA) system (Bishop et al., 1998) would thus fit to the emerging theme of TA systems playing major roles in persistence formation (Gerdes and Maisonneuve, 2012; Helaine et al., 2014). At the moment, however, it is still unclear whether increased EcnB expression is a cause or consequence of the persistent lifestyle. If the former holds true, one could speculate that STnc440 – by regulating EcnB expression – is involved in persister cell formation. (See section 5.5.3. for further information.) Future studies need to address the biological relevance of the regulation of GrxA and EcnB by STnc440 and whether they contribute to *Salmonella*'s virulence state.

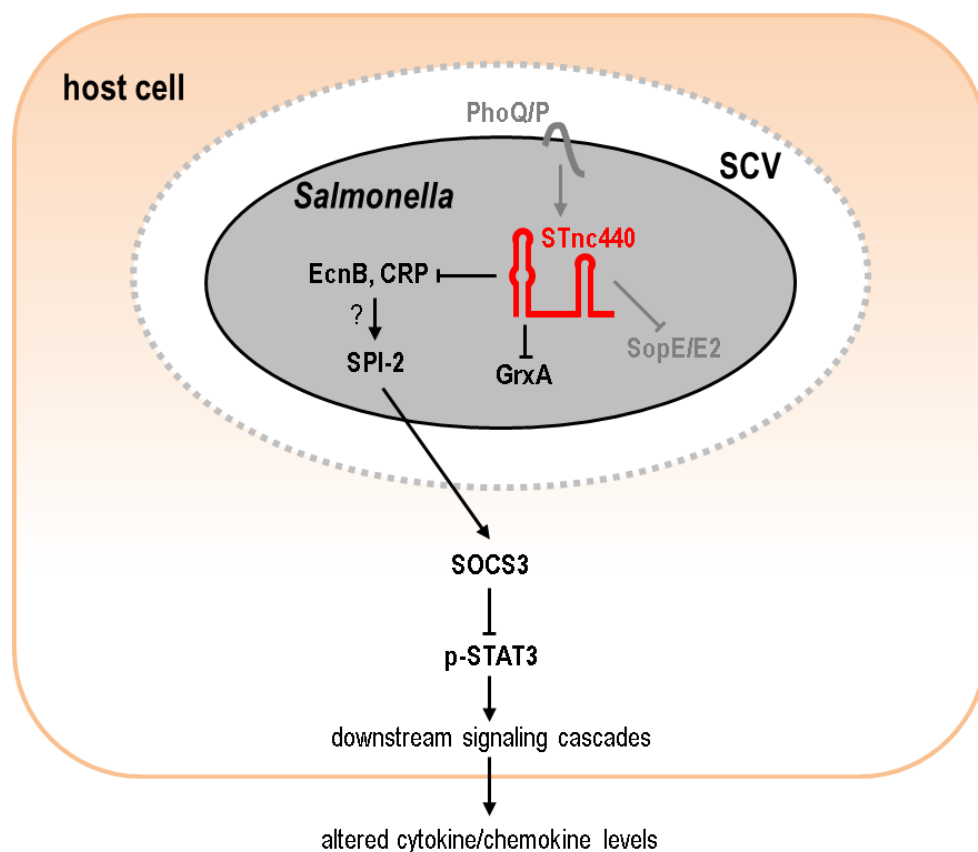


Figure 4.1/Preliminary working model for the role of STnc440 in host-pathogen interplay. A balanced expression of SPI-2 effectors that may be adjusted by STnc440 would prevent the premature induction of SOCS3, thereby enabling the activation and nuclear import of STAT3. This is accompanied by induced mRNA levels of anti-inflammatory IL-24 while pro-inflammatory IL-8 expression is appeased, presumably protecting the pathogen from an exaggerated host response. Black: findings reported in the present study; grey: Y. Chao, personal communication.

In summary, the presented data support a role for STnc440 in controlling far-reaching regulatory networks that enable *Salmonella* to adopt its specific intracellular state. In this respect, STnc440 appears reminiscent of the recently described *E. coli* sRNA McaS (Jorgensen et al., 2013) that regulates various targets via different mechanisms that converge to control biofilm formation. Similarly, several *Vibrio harveyi* sRNAs were shown to coordinate the quorum sensing response (Tu et al., 2010) and the RsaA sRNA has recently been reported to control persistence in *Staphylococcus aureus* (Romilly et al., 2014). These examples might eventually turn out to represent just the tip of the iceberg of a novel class of “lifestyle-regulating sRNAs”.

4.3 The STnc440 regulon may modulate the host response towards an appeased state

As described in the second chapter of this study, although HeLa cells are generally considered a host infection model of restricted immune pathway activity, it was capable of mounting a defined response against invading bacteria. This comprised both coding and non-coding transcript expression changes. Major differences were detected for the two sub-populations of invaded hosts and non-infected bystanders. This appears reasonable as HeLa cells neither express TLR2 nor MD2 (a critical co-receptor of TLR4) and thus, the two major cell surface PRR signaling cascades are inhibited in this cell type (Wyllie et al., 2000). Accordingly, the early stage of infection was characterized by the induction of a NF κ B response specifically in intracellularly stimulated, GFP-positive cells but not in extracellularly exposed bystanders. In fact, several of the most strongly induced mRNAs at 4 h p.i. were found to be derived from NF κ B-dependent immune genes. This indicates that intracellular PAMP sensing – e.g. via NOD1 (Buchholz and Stephens, 2008) – would be required for efficient NF κ B activation in this cell type, similar to what has been shown for THP-1 (Tukhvatulin et al., 2013). During the late infectious stage, mitochondrial genes were strongly activated – again specifically in the invaded host cell fraction. This was accompanied by the general down-regulation of the majority of other cellular transcripts. Collectively, these data might hint at a substantial degree of host cellular stress during the late stages of invasive infection. Importantly, these effects are very likely to be mitigated – and thus overseen – without the FACS-based separation step as the non-infected bystanders form by far the majority of cells in the culture. As revealed in the third chapter of this study, STnc440 expression seems to contribute significantly to certain modules of this characteristic host response, since HeLa-S3 cells containing intracellular Δ STnc440 bacteria differed from wild-type-infected cells in several aspects.

For example, host mRNAs of factors involved in JAK/STAT signaling were differentially expressed between wild-type and mutant infections. In particular, *SOCS3* induction was accelerated in Δ STnc440- compared to wild-type-infected cells. Interestingly, effectors from the SPI-2 locus have previously been suggested to influence SOCS and STAT signaling (Uchiya and Nikai, 2005, 2008). Therefore, the here-described observations would favor a model wherein balancing virulence gene expression by the STnc440 sRNA helps *Salmonella* to adjust the magnitude and/or timing kinetics of the host's immune response in a way that is ideal to establish and maintain its intracellular lifestyle (Fig. 4.1). Intriguingly, due to the fact that eventually the levels of secreted cytokines are affected in a STnc440-dependent manner, alterations will be communicated to the non-infected bystander cells and influence the entire cell population. Such a self-amplifying effect might explain how the deletion of this single bacterial sRNA could result in a fitness defect in diverse models of *Salmonella* infection.

4.4 The concept of “molecular phenotypes” of bacterial sRNAs

Data collected over the past decades has revealed an ever increasing number of bacterial sRNAs, many of which could be shown to serve key functions in orchestrating responses to changing environmental conditions and various stresses (Benjamin et al., 2010; Bobrovskyy and Vanderpool, 2014; Repoila et al., 2003; Vogel and Papenfort, 2006). For bacterial model organisms growing under diverse conditions in culture flasks, genetics and diverse biochemical assays are routinely applied to provide insight into the mechanistic aspects and cellular roles of individual sRNAs. Conversely, as infection constitutes a tight interplay of two different organisms, methods to investigate the pathogen in isolation from its host can never provide the full picture and are thus unsuited to profile virulence-associated sRNAs. Techniques, however, are lacking for the functional study of sRNAs during infection of host cells. Even though global mutant screens in animal models of infection have recently been introduced as an invaluable tool to identify bacterial genes that contribute to pathogenesis, functional analyses to uncover the underlying mechanisms have largely lagged behind. In other words, while transposon screens can address *if* a given sRNA (or any other gene) contributes to pathogenicity, methods are needed to assess *how* it would do so.

In theory, Dual RNA-seq has the potential to close this gap: Its universal character to detect and quantify all expressed transcript classes, irrespective of their kingdom affiliation, enables the simultaneous analysis of gene expression in pathogen and host. In the present study, however, a preliminary comparative Dual RNA-seq time-course of infection between wild-type and a mutant strain devoid of STnc440 did not reveal consistent changes (see above). Nevertheless, future efforts should attempt to profile STnc440-dependent gene expression changes in the

pathogen and its host by Dual RNA-seq, probably in physiologically more relevant models (e.g. primary cells or in animal models) and/or in greater depth and resolution. Furthermore, as infection is a highly variable biological process it is desirable to further decrease the material constraints for cDNA library construction to allow gene expression to be monitored in more defined sub-populations of infected cells (see sections 5.5.3 and 5.5.4). This way, it might eventually become possible to resolve the “footprint” of a given bacterial sRNA (or any other virulence factor) on both the pathogen and host transcriptomes during infection.

Traditional techniques to profile the phenotypes of individual sRNA mutants (e.g. c.f.u. assays or host cell death measurements) can be considered end-point determinations as the net outcome of infection is assessed. Dual RNA-seq would in principle allow effects to be appreciated much earlier as the transcriptome rapidly adapts to changing conditions and therefore mirrors the physiological state of the cells at any given time point after infection. Such a “molecular phenotype” might or might not eventually lead to a “macroscopic phenotype” (i.e. a virulence defect). In the long run Dual RNA-seq is proposed as a method to complement macroscopic with molecular phenotypes for virulence genes during bacterial infections of eukaryotic cells, to provide a new angle on host-pathogen interaction studies.

4.5 Future applications

The HeLa-S3 infection model proved suitable for the purpose of evaluating the general feasibility of Dual RNA-seq and revealed new aspects of *Salmonella* infection. Future studies should aim at transferring the method to model systems closer to physiological conditions of the infection with *Salmonella* and other pathogens. In the following, a selection of scientific questions are listed, for which to answer Dual RNA-seq is envisaged to be helpful in the future.

4.5.1 Molecular phenotyping of deletion mutants of entire sRNA regulons during infection

STnc440 was one of the most induced *Salmonella* genes during infection of both epithelial and monocytic host cells and gave rise to virulence defects when inhibited. In contrast, several further host-induced sRNAs identified in the present work did not produce a positive hit in global transposon screens for *Salmonella* virulence genes (Barquist et al., 2013; Chaudhuri et al., 2013; Santiviago et al., 2009). This means that the mere intracellular induction of a given sRNA cannot be considered strict evidence for it giving rise to a virulence phenotype when deleted. To some extent this may be due to functional redundancy, as several sRNAs have been shown to regulate the same mRNA target. Additionally sRNAs generally do not function as on/off switches but rather as timers or fine-tuners of target gene expression. The present Dual RNA-seq data provides a resource for the identification of co-regulated sRNAs and the prediction of their

corresponding activator. Deleting functionally related sRNA pairs (see Appendix) or entire sRNA regulons, rather than individual sRNAs alone, combined with profiling the *in vivo* performance of the resulting mutants could inform about the joint contribution of functionally redundant or co-regulated sRNAs to bacterial virulence.

For example, several sRNAs are known to be induced together with major virulence programs under infection conditions in *Salmonella* or pathogenic *E. coli*. Previous work on IsrM demonstrated that this sRNA is activated under conditions of low pH and increased osmolarity and to regulate the expression of *hlyE*, encoding a negative regulator of SPI-1 (Gong et al., 2011). Furthermore, there are currently three confirmed HilD/HilA-responsive sRNAs, InvR (Pfeiffer et al., 2007), DapZ (Chao et al., 2012) and IsrJ (Padalon-Brauch et al., 2008), and two PhoP-dependent sRNAs, MgrR (Moon et al., 2013) and STnc440 (Y. Chao, personal communication). However, intracellular sRNA expression kinetics derived from temporal Dual RNA-seq profiling of wild-type *Salmonella* revealed further sRNAs to have similar expression patterns and suggest that their production might be also HilD- or PhoP-driven. An assessment of multi-mutant strains devoid of these putative regulons of SPI-1- or SPI-2-associated sRNAs might help solve the current controversy concerning the fact that Δhfq mutants of many pathogens are highly attenuated while individual sRNA deletion strains display only modest (if any) virulence phenotypes (Chao and Vogel, 2010).

4.5.2 Towards *in vivo* Dual RNA-seq

As was shown for infected BMDMs in the present work, Dual RNA-seq can be applied to isolated and *in vitro*-infected primary cells. However, primary cells appeared to be more sensitive to RNAlater fixation and/or the mechanical stress applied during cell sorting than immortalized cell types. Therefore, for the design of larger-scale experiments in the future, the current protocol will have to be optimized for primary cells. Future attempts should also include the transfer of the Dual RNA-seq technique to three-dimensional tissue models (e.g. Pusch et al., 2011) or to animal models such as the streptomycin-treated mouse (Barthel et al., 2003).

4.5.3 Intra-species pathogen heterogeneity: transcriptomics for bacterial persisters and their hosts

Genetically identical cells do not necessarily behave alike. Instead, many bacterial species are known to display extensive phenotypic variation, e.g. with respect to their division rate or susceptibility to antibiotics. A prominent example is bacterial persistence (Balaban et al., 2013). Persisters are defined as multidrug-tolerant bacterial cells causing relapsing, chronic infections. Moreover, persistence usually comprises a non-replicating, dormant cell state. Recently novel tools have emerged that foster the study of bacterial persisters (Helaine and Kugelberg, 2014). For instance, fluorescence-based techniques have been successfully used both *in vitro* and *in vivo*

to demonstrate how host tissues and sub-cellular micro-environments diversify *Salmonella* phenotypic heterogeneity with consequences for disease progression and the development of therapeutics to eradicate persisting individuals (Claudi et al., 2014; Helaine et al., 2014). The combination of these techniques with Dual RNA-seq could permit unprecedented insight into the molecular mechanisms underlying the persistence phenotype as well as into the host response to persistent pathogens.

4.5.4 Single-cell Dual RNA-seq

Not only bacterial but also eukaryotic cells occasionally vary dramatically with respect to their behavior despite an identical genetic background (examples are reviewed in Saliba et al., 2014). This cell-to-cell variability is known to be reflected in the transcriptomic composition between individuals. As infection comprises two organisms, that both may display a certain extent of heterogeneity, the total variability between two infected cells is expected to exceed the variation between two uninfected host cells or between extracellular bacteria. In other words, variability potentiates with the number of different organisms involved in a biological process.

In recent years, ever-increasing sensitivity of research instruments together with the automation of many experimental steps from single-cell isolation to sample preparation and data analysis have facilitated single-cell studies, including single-cell transcriptomics. The major challenge the field of single-cell RNA-seq currently faces is to broaden the range of detectable transcripts from polyadenylated-only to *all* types of cellular RNA classes. In parallel, sensitivity will have to increase to make the method eventually compatible with the minute RNA amounts of a single prokaryotic cell. Once these prerequisites have been attained, Dual RNA-seq on single infected cells should become feasible and provide unprecedented resolution to infection biology.

4.5.5 Bacterial/viral co-infections

Viruses hijack the transcription machinery of their host, meaning that a virus-infected cell will produce only a single transcriptome (even if the resulting reads from a RNA-seq experiment map to two different genomes). For example, in line with viral mRNAs being transcribed by the host machinery, they become capped and polyadenylated. In addition, viral transcripts will usually be produced at very high amounts. Consequently, RNA-seq-based transcriptome analysis of virus-infected host cells is straight-forward (Lisnic et al., 2013).

It is known that infection by a pathogen may heighten the likelihood of a secondary infection with a different pathogenic species. Such co-infections occur frequently between several bacterial pathogens and certain viruses, such as *Streptococcus* spp. and influenza virus (Chertow and Memoli, 2013), *Trichomonas vaginalis* and human immunodeficiency virus (Kissinger and Adamski, 2013) or *Chlamydia trachomatis* and human herpes virus (Prusty et al., 2013). The

genome-independent nature of RNA-seq again bears great potential to analyze gene expression changes of the host together with those of its multiple pathogens.

4.6 Conclusions and perspectives

Compared to traditional methods, RNA-seq provides unprecedented access to the transcriptomes of potential hosts or pathogens, it improves the annotation of genomes and enables the detection of novel genes. This has caused a shift from hypothesis-based to unbiased, discovery-driven approaches, thereby increasing the likelihood of identifying novel principles. The described Dual RNA-seq technique comprises many of the advantages associated with classical RNA-seq but profiles gene expression in host and pathogen simultaneously while they physically interact. It can therefore provide new insights into the infection process that could not be obtained before by sequencing the individual partners in isolation. The present work focuses largely on the contribution of *Salmonella* sRNAs to the adoption of the pathogen's intracellular state and the contextual manipulation of its host. However, similar analyses appear feasible for other bacterial virulence genes or host factors that define susceptibility or contribute to defense mechanisms towards bacterial infection.

Furthermore, future research may shift from the analysis of 'average' cellular responses to the detection of more specific expression programs in defined sub-populations of host cells challenged with pathogens (e.g. bystanders, invaded hosts containing bacterial loads in varying amounts). Likewise, there is variability within the bacterial population, e.g. with respect to the pathogenicity of individual members (i.e. replicating vs. persisting bacteria). The synchronized detection of host and pathogen expression by Dual RNA-seq can be applied to assess the role of stochasticity and cell type-specificity for the development of such phenotypes; features that would not be addressable by sequencing the transcriptome of either host or pathogen alone.

Ongoing advances in 'third-generation' sequencing methods are expected to further boost multi-transcriptomics approaches. In particular, recent improvements in nanopore sequencing (Loman and Quinlan, 2014; Mattei, 2014; Rosenstein, 2014; Ying et al., 2014) appear highly promising, as this technology is potentially compatible with direct RNA sequencing (i.e. without a cDNA intermediate) and without read length limitation. The expected increase in throughput and sensitivity might enable scientists to resolve gene expression events during infection in finer detail by scaling down the number of cells – potentially down to the single-cell level.

5. Material and methods

5.1 Material

5.1.1 Technical instruments

Table 5.1/List of instruments and devices used.

Instrument/device	Manufacturer
analytical balances TE64, TE601	Sartorius
Bio-Link BLX 254 UV-Crosslinker	PeqLab
centrifuge Eppendorf 5415R	Eppendorf
centrifuge Eppendorf 5424	Eppendorf
centrifuge Hereaus Multifuge X3R	Thermo Scientific
eraser for imaging screens	Molecular Dynamics
FACSAria III	BD Biosciences
FACSCalibur	BD Biosciences
gel documentation system Gel iX Imager	INTAS UV Systeme
heat block Thermomixer comfort	Eppendorf
horizontal electrophoresis system Perfect Blue Mini S, M, L	PeqLab
hybridization oven HP-1000	UVP
imaging system Image Quant LAS 4000	GE
incubator for bacterial plates	Memmert
incubator for cell culture flasks HERAccl 150i	Thermo Scientific
Infinite F200 PRO plate reader	TECAN
Innova 44 shaker	New Brunswick Scientific
LEICA SP5 confocal microscope	Leica
light microscope Eclipse TS100	Nikon
MACSQuant Analyzer	Miltenyi Biotec
MicroPulser electroporator	BioRad
MiSeq	Illumina
Multiskan Ascent	Thermo Electron Corporation
PCR cycler MJ Mini	BioRad
phosphoimager Typhoon FLA 7000	GE
photometer Ultraspec 10 Cell Density Meter	Amersham Biosciences
power supplies peqPOWER E300	PeqLab
Qubit 2.0 Fluorometer	Life Technologies
real time PCR cycler CFX96 Real-Time System	BioRad
Refrigerated Incubator Shaker C24KC	PeqLab
Safe 2020 cell culture hood	Thermo Scientific

Table 5.1/continued.

Instrument/device	Manufacturer
semi-dry electroblotter Perfect Blue SEDEC M	PeqLab
shaking water bath incubator 1092	GFL
spectrophotometer NanoDrop 2000	Thermo Scientific
VacuSafe pump	IBS Integra Biosciences
vertical electrophoresis system Perfect Blue Twin S, ExW S, L	PeqLab
Victor3 1420 Multilabel Counter	PerkinElmer
Vortex-Genie 2	Scientific Industries
water bath HAAKE A10	Thermo Scientific

5.1.2 Glass/plastic ware and consumables

Table 5.2/List of lab ware used.

Labware	Manufacturer
12-well plates	Corning
24-well plates	Corning
50 mL syringe	BD Plastipak
6-well plates	Corning
96-well microtiter plates	Nunc
cell scraper 25 cm	Sarstedt
Cellstar serological pipets 5 mL, 10 mL, 25 mL (plastic)	Greiner bio-one
cover slips	Hartenstein
electroporation cuvettes (2 mm)	Cell projects
G-25, G-50 MicroSpin columns	GE Healthcare
Gilson pipets 10 μ L, 20 μ L, 200 μ L, 1,000 μ L	Gilson
glass beads (0.1 mm)	Roth
glass bottles	Schott
glass test tubes and lids	Roth
glass tips (for pump)	Geyer
Hard-shell PCR plates 96-well WHT/WHT	BioRad
Hybond-XL Membrane for nucleic acid transfer	GE Healthcare
imaging cassettes	Fujifilm
imaging screens	Fujifilm
inoculation loops (10 μ L)	VWR
L-shape bacteriology loops	VWR
MACS pre-separation filters (30 μ m pore size)	Miltenyi Biotec
Multidispenser combitips (5 mL)	Eppendorf
Multidispenser Repeater Plus	Eppendorf
Neubauer counting chamber	HBG Henneberg-Sander

Table 5.2*/continued.*

Labware	Manufacturer
object slides	Hartenstein
PCR tubes (8 x 0.5 mL stripes)	Thermo Scientific
petri dishes	Corning
phase-lock gel (PLG) tubes 1.5 mL, 2 mL	5 PRIME
Pipetboy acu-jet pro	BRAND
Pipetman P10, P20, P200, P1000 pipet tips	Sarstedt
PolyScreen PVDF transfer membrane	PerkinElmer
reaction tubes 15 mL, 50 mL	Sarstedt
safe-lock tubes 1.5 mL, 2 mL	Eppendorf
spectrophotometer cuvettes	Sarstedt
Sterile filters (0.2 µm pore size)	Sarstedt
T-150 flasks	Corning
T-25 flasks	Corning
T-75 flasks	Corning
TC Microwell 96F	Thermo Scientific

5.1.3 Chemicals, reagents, proteins and size markers

Table 5.3*/List of chemicals, reagents and proteins used.*

Chemicals/reagents/proteins/size markers	Manufacturer
6x DNA loading buffer	Fermentas
Albumin Fraktion V	Roth
ampicillin sodium salt	Roth
Annexin V-APC	BD Pharmingen
chloramphenicol	Roth
D(+)-glucose	Merck
dimethyl sulfoxide (DMSO)	Roth
DMEM	Gibco
EDTA	Roth
ethanol	Roth
ethanol (absolute)	Merck
Gel loading buffer II	Ambion
GeneRuler 1 kb DNA Ladder	Fermentas
gentamicin sulfate salt	Sigma
GlycoBlue	Ambion
GlycoBlue	Ambion
H ₂ O ₂ (30%)	AppliChem
Igpal	Sigma
isopropanol	Roth

Table 5.3/continued.

Chemicals/reagents/proteins/size markers	Manufacturer
kanamycin sulfate	Roth
L(+)-arabinose	Roth
methanol	Roth
milk powder	Roth
PBS	Gibco
Prestained Protein Marker Broad Range	New England Biolabs
propidium iodide	Sigma
pUC Mix Marker, 8	Fermentas
random hexamers	Invitrogen
RedSafe	ChemBio
RedSafe	ABC Scientific
RNA Ladder High Range	Fermentas
RNA Ladder Low Range	Fermentas
RNA _{later}	Qiagen
Roti-Aqua P/C/I	Roth
Roti-Free	Roth
Roti-Hybi-Quick	Roth
Rotiphorese gel 40 (19:1)	Roth
Rotiphorese gel 40 (37.5:1)	Roth
RPMI	Gibco
SUPERaseIN RNase Inhibitor	Ambion
tetracycline	Roth
Triton X-100	Sigma
TRIzol reagent	Invitrogen
Western Lightning chemiluminescence reagent	PerkinElmer

5.1.4 Commercial kits

Table 5.4/Commercial kits used.

Kit	Manufacturer
CytoTox 96(R) Non-Radioactive Cytotoxicity Assay	Promega
Human CXCL8/IL-8 DuoSet	R&D Systems
MAXIsript T7 Kit	Ambion
MEGAscript T7 Kit	Ambion
mirVana miRNA isolation kit	Life technologies
NucleoSpin Extract II	Macherey-Nagel
NucleoSpin Plasmid QuickPure	Macherey-Nagel
Power SYBR Green RNA-to-CT 1-Step Kit	Life technologies
P-STAT sampler kit	Cell Signaling Technology

Table 5.4*/continued.*

Kit	Manufacturer
Rediprime II DNA Labeling System	GE Healthcare
SV40 Total RNA Isolation System	Promega
TOPO TA Cloning Kit	Invitrogen

5.1.5 Enzymes

Table 5.5*/List of enzymes used.*

Enzyme	Manufacturer
Calf Intestine Alkaline Phosphatase (CIP, 10 U/μL)	New England Biolabs
Deoxyribonuclease I (DNase I, 1 U/μL)	Fermentas
<i>DpnI</i> (20 U/μL)	New England Biolabs
<i>Pfu</i> DNA Polymerase (2.5 U/μL)	Fermentas
Phusion High-Fidelity DNA Polymerase (2 U/μL)	New England Biolabs
Shrimp Alkaline Phosphatase (SAP, 1 U/μL)	Fermentas
SuperScript II Reverse Transcriptase	Invitrogen
SuperScript III Reverse Transcriptase	Invitrogen
T4 DNA Ligase (5 U/μL)	Fermentas
<i>Taq</i> DNA polymerase (5 U/μL)	New England Biolabs
Tobacco Acid Pyrophosphatase (TAP, 10 U/μL)	Epicentre
Trypsin-EDTA (0.25%), phenol red	Gibco
Various restriction enzymes from New England Biolabs or Fermentas	

5.1.6 Antibodies

Table 5.6*/List of primary and secondary antibodies and -sera used.*

Antibody/antiserum	Source	Dilution	Provider
anti-FLAG	mouse	(1:1,000 in 3% BSA)	Sigma
anti-GroEL	rabbit	(1:10,000 in 3% BSA)	Sigma
anti-CRP	rabbit	(1:500 in 3% BSA)	Hiroji Aiba (Nagoya, Japan)
anti-phospho-Stat3 (Tyr705)	rabbit	(1:1,000 in 3% BSA)	Cell Signaling Technology
anti-phospho-Stat3 (Ser727)	rabbit	(1:1,000 in 3% BSA)	Cell Signaling Technology
anti-tubulin	mouse	(1:10,000 in 3% BSA)	Ana Eulalio (Würzburg, Germany)
anti-mouse	goat	(1:10,000 in 3% BSA)	Thermo Scientific
anti-rabbit	goat	(1:10,000 in 3% BSA)	Thermo Scientific

5.1.7 Synthetic oligonucleotides

Table 5.7/List of synthetic oligonucleotides used.

name	target	purpose	nucleotide sequence
<i>Salmonella</i> -directed oligonucleotides			
JVO-1117	<i>Salmonella rfaH</i>	qRT-PCR sense primer	5'-TCAGCCATTTTGTGCGCTT-3'
JVO-1118	<i>Salmonella rfaH</i>	qRT-PCR antisense primer	5'-TTCAGATCGACAACGCCTT-3'
AE-1	<i>gfp</i>	qRT-PCR sense primer	5'-ATGCTTTTCCCCTTATCCGG-3'
AE-2	<i>gfp</i>	qRT-PCR antisense primer	5'-GCGTCTTGTAGTTCCTGTCATC-3'
JVO-4621	<i>Salmonella</i> STnc440	qRT-PCR sense primer	5'-AGTAACGGATTACTTTGTGGT-3'
JVO-9040	<i>Salmonella</i> STnc440	qRT-PCR antisense primer	5'-ACTACGCTGCCGCAATAT-3'
JVO-9804	<i>Salmonella</i> RyhB	qRT-PCR sense primer	5'-GCATCCAGGGGAACCC-3'
JVO-9805	<i>Salmonella</i> RyhB	qRT-PCR antisense primer	5'-CAGCAAAGCTGGCCAA-3'
JVO-8216	<i>Salmonella</i> RybB	qRT-PCR sense primer	5'-GCCACTGCTTTTCTTTGATGT-3'
JVO-8217	<i>Salmonella</i> RybB	qRT-PCR antisense primer	5'-AACCCACCAACCTTGAACC-3'
JVO-10198	<i>Salmonella</i> STnc510	qRT-PCR sense primer	5'-TAAATCGAGCGGGAATAAG-3'
JVO-10199	<i>Salmonella</i> STnc510	qRT-PCR antisense primer	5'-GCTTCACAGTCCTCACGTGAC-3'
JVO-322	<i>Salmonella</i> 5S rRNA	Northern probe	5'-CTACGGCGTTTCACTTCTGAGTTC-3'
JVO-9982	<i>Salmonella</i> tRNA-fMet	Northern probe	5'-GACGAGCTACCAGGCTGCT-3'
JVO-10085	<i>Salmonella</i> tRNA-Ser	Northern probe	5'-TCAAGACCGGTGCCTTCAA-3'
JVO-2408	<i>Salmonella</i> STnc440	Northern probe	5'-TATGAGGAGGACAATTACCG-3'
JVO-10254	<i>Salmonella</i> sseA	qRT-PCR sense primer	5'-CTTGCTGAAAGGCGAGAGAG-3'
JVO-10255	<i>Salmonella</i> sseA	qRT-PCR antisense primer	5'-TTCTTTAAATCCTTCTCGGCC-3'
JVO-10250	<i>Salmonella</i> sseC	qRT-PCR sense primer	5'-GAGGTGCAAAATGCATTACGTA-3'
JVO-10251	<i>Salmonella</i> sseC	qRT-PCR antisense primer	5'-CTTTACGCGCTTTATCCTCCT-3'
JVO-11545	<i>Salmonella</i> steC	qRT-PCR sense primer	5'-GACGGATGCATCACAACCAG-3'
JVO-11546	<i>Salmonella</i> steC	qRT-PCR antisense primer	5'-CATGTAACCGATAGACTGTCCCA-3'
JVO-10290	<i>Salmonella</i> ssrB	qRT-PCR sense primer	5'-GGAAGCTATCCTGGCTGAATTA-3'
JVO-10291	<i>Salmonella</i> ssrB	qRT-PCR antisense primer	5'-ATTCATCCGGTGTGTTTCG-3'
JVO-10398	<i>Salmonella</i> phoP	qRT-PCR sense primer	5'-AAGAGGTAATGGCGGTATG-3'
JVO-10399	<i>Salmonella</i> phoP	qRT-PCR antisense primer	5'-GATGACCTCTTCATTGACGGA-3'
JVO-5528	<i>Salmonella</i> hiiD	qRT-PCR sense primer	5'-GTACGTCAACGCTCAAACG-3'
JVO-566	<i>Salmonella</i> hiiD	qRT-PCR antisense primer	5'-GAAACATTGAATGAAGTAGGACG-3'
JVO-9805	<i>Salmonella</i> RyhB	Northern probe	5'-CAGCAAAGCTGGCCAA-3'
JVO-9807	<i>Salmonella</i> lsrE	Northern probe	5'-AGGCGGGCAAATAACTG-3'
JVO-1001	<i>Salmonella</i> OmrA	Northern probe	5'-GAGACAGGTACGAAGAGCGTAC-3'
JVO-1000	<i>Salmonella</i> OmrB	Northern probe	5'-TGTGATCAACAATGACGTTGA-3'
JVO-12151	<i>Salmonella</i> YrlA	Northern probe	5'-ATTGAGCTACAGAGCCGGG-3'
JVO-12152	<i>Salmonella</i> YrlB	Northern probe	5'-TCATAGCCGGGATGATACG-3'
JVO-10705	<i>Salmonella</i> grxA	sense oligo to add 3xFLAG to <i>grxA</i> in SL1344	5'-TTTTGAAGCATGGGCAAAAGAAATCTGAATCTTTTCGCC GACTACAAAGACCATGACGG-3'
JVO-10706	<i>Salmonella</i> grxA	antisense oligo to add 3xFLAG to <i>grxA</i> and antisense deletion primer of <i>grxA</i> in SL1344	5'-CGCCGATAGCGCCCTCCGTCAGAGGGCGTTGCCAGAGC AACCATATGAATATCCTCCTTAG-3'
JVO-11247	<i>Salmonella</i> grxA	sense deletion primer of <i>grxA</i> in SL1344 (-40 to -1 upstream of AUG)	5'-ACCTGTTCCCGTCTGGCTTTATGTTTCGAGGAGAGATATG GACTACAAAGACCATGACGG-3'
JVO-11657	<i>Salmonella</i> steC	sense oligo to add 3xFLAG to <i>steC</i> in SL1344	5'-GACTCTTTGTGGCTAAGGTATTAAGGATGAATAAAAAAA GACTACAAAGACCATGACGG-3'
JVO-11658	<i>Salmonella</i> steC	antisense oligo to add 3xFLAG to <i>steC</i> in SL1344	5'-TGTGCCCGGCGATTGCGAGAAAAGAACGGAATAAAT GCCATATGAATATCCTCCTTA-3'
JVO-11657	<i>Salmonella</i> steC	sense oligo to add 3xFLAG to <i>steC</i> in SL1344	5'-GACTCTTTGTGGCTAAGGTATTAAGGATGAATAAAAAAA GACTACAAAGACCATGACGG-3'

Table 5.7/continued.

name	target	purpose	nucleotide sequence
JVO-11658	<i>Salmonella steC</i>	antisense oligo to add 3xFLAG to <i>steC</i> in SL1344	5'-TGTGCCCCCGGCGATTGCGAGAAAAGAACGGAACATAATGCCATATGAATATCCTCCTTA-3'
JVO-10703	<i>Salmonella grxA</i>	sense oligo to generate <i>grxA::gfp</i> fusion	5'-GTTTATGCATGTTGCGCGTCTGGCTTTA-3'
JVO-10704	<i>Salmonella grxA</i>	antisense oligo to generate <i>grxA::gfp</i> fusion	5'-GTTTTGCTAGCATCGTCGCGTTCTTTGCTC-3'
JVO-11991	<i>Salmonella grxA</i>	sense mutagenesis primer to generate <i>grxA**</i> construct	5'-TTCGAGGAGAGATATGTTAGTGTATTTTGGCCGTCCAGG-3'
JVO-11759	<i>Salmonella grxA</i>	antisense primer to generate <i>grxA**</i> construct	5'-TCTCCTCGAACATAAAGCCAGAC-3'
JVO-4224	<i>Salmonella crp</i>	sense oligo to generate <i>crp::gfp</i> fusion (1°TSS)	5'-GTTTTTTATGCATTTCCAGGTAGCGGGAAGT-3'
JVO-4225	<i>Salmonella crp</i>	antisense oligo to generate <i>crp::gfp</i> fusion (1°TSS)	5'-GTTTTTTCTAGAAAACCGCAAACAGACCCG-3'
JVO-5850	<i>Salmonella crp</i>	sense oligo to generate <i>crp::gfp</i> fusion (2°TSS)	5'-GTTTTTTATGCATGATGTACAGTAATACATTGACGTAC-3'
JVO-5851	<i>Salmonella crp</i>	antisense oligo to generate <i>crp::gfp</i> fusion (2°TSS)	5'-GTTTTTTGCTAGCACGGGTGCCGTAGACGAC-3'
JVO-4621	<i>Salmonella</i> STnc440	sense oligo to clone STnc440	5'-p-AGTAACGGATTACTTTGTGGTG-3'
JVO-4724	<i>Salmonella</i> STnc440	antisense oligo to clone STnc440	5'-GTTTTTTCTAGATTGTCTGTTAATTATTACAGAGAGAG-3'
JVO-5853	<i>Salmonella</i> STnc440	sense oligo to clone pOWN-STnc440	5'-GTTTTTTCTAGACTGTCACAGTATGCTAAACAG-3'
JVO-5473	<i>Salmonella</i> STnc440	sense oligo to introduce STnc440* mutation	5'-CGTAACCCTAATTGTCTCTCATATTTGC-3'
JVO-5474	<i>Salmonella</i> STnc440	antisense oligo to introduce STnc440* mutation	5'-CAATTAGGGTTACGCTACACCACAAAGTAAT-3'
JVO-8695	<i>Salmonella ssaG</i>	sense oligo to clone <i>ssaG</i> promoter	5'-GTTTTGACGTCGAGTGGTAGTTTGGACTAC-3'
JVO-8696	<i>Salmonella ssaG</i>	antisense oligo to clone <i>ssaG</i> promoter	5'-GTTTTGCTAGCTTTAATCATCGATTCTGG-3'
JVO-4726	<i>Salmonella</i> STnc440	sense oligo to disrupt STnc440	5'-AATCTATTTTACTGTCACAGTATGCTAAACAGAACAGGTGATAGGCTGGAGCTGCTTC-3'
JVO-8070	<i>Salmonella</i> STnc440	antisense oligo to disrupt STnc440	5'-GAGTTAATTTATAAAAAAGCGGCAGACTACGCTGCCCGGTCCATATGAATATCCTCC-3'
JVO-8739	<i>Salmonella</i> STnc440	sense oligo to introduce trans-complemented <i>STnc440::Cm</i> in <i>put</i> locus	5'-GCAGGTACATTTAACGCGGTTGCACAAGTTGCAACATGCCGTCTACTGTCCCTAGTG-3'
JVO-8740	<i>Salmonella</i> STnc440	antisense oligo to introduce trans-complemented <i>STnc440::Cm</i> in <i>put</i> locus	5'-GACCCGGATAGTAATTTGCCCGCCAGATGATAAATCGCGATGTCCTACTCAGGAGAG-3'
JVO-5796	<i>Salmonella</i> STnc440	sense oligo to exchange <i>Cm</i> to <i>dhfr</i>	5'-GTTCTGAGGTCATTACTGGATCTATCAACAGAGTCCAA GTACACATCTCAACCATCAT-3'
JVO-5797	<i>Salmonella</i> STnc440	sense oligo to exchange <i>Cm</i> to <i>dhfr</i>	5'-TATCTGGCGAAAATGAGACGTTGATCGGCACGTAAGAGGTTTATACACATCTCAACCCT-3'
JVO-9313	<i>Salmonella</i> STnc440	sense oligo to introduce trans-complemented <i>STnc440::dhfr</i> in <i>trg</i> locus	5'-TTAACGCTTATATGTCCAGAGCGGTGAAAACACCCGCCGACTTACTGTCCCTAGTGCTTG-3'
JVO-9314	<i>Salmonella</i> STnc440	antisense oligo to introduce trans-complemented <i>STnc440::dhfr</i> in <i>trg</i> locus	5'-CGTTCAGCGCCAGGATGTTGGTCTGGAAAGCGATACTGTTCGGATTGTCTACTCAGGA-3'
human-directed oligonucleotides			
JVO-7673	human U6 snRNA	qRT-PCR sense primer	5'-GCTTCGGCAGCACATATACTAAAAT-3'
JVO-7672	human U6 snRNA	qRT-PCR antisense primer	5'-ATATGGAACGCTTCACGAATTTG-3'
JVO-5409	human β -actin	qRT-PCR sense primer	5'-CCTGTACGCCAACACAGTGC-3'
JVO-5410	human β -actin	qRT-PCR antisense primer	5'-ATACTCTGCTTGCTGATCC-3'
JVO-12865	human SOCS1	qRT-PCR sense primer	5'-GACTGCCTCTCGAGCTGC-3'
JVO-12866	human SOCS1	qRT-PCR antisense primer	5'-CTGGAAGGGGAAGGAGCTC-3'
JVO-11432	human SOCS2	qRT-PCR sense primer	5'-GCCAAAGAGAAATTAAGAGGCA-3'
JVO-11433	human SOCS2	qRT-PCR antisense primer	5'-GAAGATTAGTTGGTCCAGCTGATG-3'
JVO-12867	human SOCS3	qRT-PCR sense primer	5'-TGGTCACCCACAGCAAGTTT-3'
JVO-12868	human SOCS3	qRT-PCR antisense primer	5'-GGCACTGAGCAGCAGGTTTC-3'
JVO-10352	human <i>IL24</i>	qRT-PCR sense primer	5'-TCCCTTGCTGGGTTTAC-3'
JVO-10353	human <i>IL24</i>	qRT-PCR antisense primer	5'-CCAGAAGGCTTCCACAGT-3'
JVO-8832	human <i>IL8</i>	qRT-PCR sense primer	5'-CTGGCCGTGGCTCTCTTG-3'
JVO-8833	human <i>IL8</i>	qRT-PCR antisense primer	5'-CCTTGGCAAACTGCACCTT-3'

5.1.8 Plasmids

Table 5.8/List of plasmids used.

trivial name	stock number	relevant insert	comment	parental plasmid	resistance marker	reference
pBAD-ctrl.	pKP8-35		empty control plasmid		Amp	
pBAD-STnc440	pYC5-34	STnc440	wild-type STnc440 under arabinose-inducible promoter	pBAD	Amp	this study
pBAD-STnc440*	pYC60	STnc440*	double point-mutated STnc440 under arabinose-inducible promoter	pBAD	Amp	this study
FLP helper plasmid	pCP20		used to heal genetically modified SL1344 derivatives			Datsenko and Wanner, 2000
pOWN-ctrl.	pJV300		empty control plasmid	pZE12	Amp	this study
pOWN-STnc440	pYC55	STnc440	wild-type STnc440 under native promoter	pZE12	Amp	this study
pXG-10			empty control plasmid		Cm	
<i>grxA::gfp</i>	pYC213	<i>grxA::gfp</i>	<i>grxA</i> -30aa translational GFP fusion	pXG-10	Cm	this study
<i>crp</i> (1°TSS):: <i>gfp</i>	pYC53	<i>crp::gfp</i> (1°TSS)	<i>crp</i> -10aa translational GFP fusion	pXG-10	Cm	this study
<i>crp</i> (2°TSS):: <i>gfp</i>	pYC51	<i>crp::gfp</i> (2°TSS)	<i>crp</i> -50aa translational GFP fusion fusion	pXG-10	Cm	this study
pXG-0			negative control vector		Cm	
pXG-1		<i>gfp</i>	positive control vector		Cm	
<i>ssaG::gfp</i>	pYC104	<i>ssaG</i> promoter	<i>ssaG</i> promoter fused to <i>gfp</i>	pAS093	Cm	this study

5.1.9 Bacterial strains

Table 5.9/List of *Salmonella* strains used.

trivial name	stock name	resistance marker	comments	reference
WT	JVS-3858	Str, Cm	constitutively GFP-expressing SL1344 strain (GFP integrated in the chromosome; <i>ptet-GFP</i>)	Papenfort et al., 2009
Δ STnc440	JVS-10038	Str, Cm	deletion of STnc440 (terminator was kept) in the <i>ptet-GFP</i> background	this study
Δ STnc440+	JVS-10039	Str, Cm, dhfr	<i>trans</i> -complementation of STnc440 in the <i>ptet-GFP</i> background (complementation in <i>trg</i> locus)	this study
<i>grxA::3xflag</i>	JVS-11005	Str	epitope-tagged endogenous <i>grxA</i> on the chromosome in the wild-type background	this study
Δ STnc440/ <i>grxA::3xflag</i>	JVS-11006	Str	epitope-tagged endogenous <i>grxA</i> on the chromosome in the Δ STnc440 background	this study
Δ STnc440/ Δ <i>grxA</i>	JVS-11019	Str	deletions of STnc440 and <i>grxA</i> from the chromosome	this study
<i>ecnB::3xflag</i>	JVS-11071	Km, Str	epitope-tagged endogenous <i>ecnB</i> on the chromosome in the wild-type background	this study
Δ STnc440/ <i>ecnB::3xflag</i>	JVS-11072	Km, Str	epitope-tagged endogenous <i>ecnB</i> on the chromosome in the Δ STnc440 background	this study
Δ <i>grxA</i>	JVS-11074	Str	deletion of <i>grxA</i>	this study
Δ STnc440/ Δ <i>crp</i>	JVS-11011	Str, Tet	deletion of STnc440 and Tn10 insertion into <i>crp</i> gene locus	this study
Δ <i>crp</i>	JVS-1626	Str, Tet	Tn10 insertion into <i>crp</i> gene locus	P22 lysate was obtained from M. Teplitski (University of Florida); originally generated by J. Roth
<i>sopE::3xflag/steC::3xflag</i>	JVS-10981	Str, Kan	epitope-tagged endogenous <i>sopE</i> and <i>steC</i> on the chromosome in the wild-type background	this study

Table 5.9*/continued.*

trivial name	stock name	resistance marker	comments	reference
Δ STnc440_ <i>sopE</i> ::3 <i>xflag</i> / <i>steC</i> ::3 <i>xflag</i>	JVS-11009	Str	epitope-tagged endogenous <i>sopE</i> and <i>steC</i> on the chromosome in the Δ STnc440 background	this study
Δ STnc440+_ <i>sopE</i> ::3 <i>xflag</i> / <i>steC</i> ::3 <i>xflag</i>	JVS-10983	Str, Kan	epitope-tagged endogenous <i>sopE</i> and <i>steC</i> on the chromosome in the Δ STnc440+ background	this study
Δ <i>ryhB</i> / Δ <i>isrE</i>	JVS-10996	Str, Kan	deletions of <i>ryhB</i> and <i>isrE</i> from the chromosome in the <i>ptet-GFP</i> background	this study
Δ <i>omrA</i> / Δ <i>omrB</i>	JVS-10792	Str, Cm	deletions of <i>omrA</i> and <i>omrB</i> from the chromosome in the <i>ptet-GFP</i> background	this study
Δ <i>yrjA</i> / Δ <i>yrjB</i>	JVS-10977	Str, Cm	deletions of <i>yrjA</i> and <i>yrjB</i> from the chromosome in the <i>ptet-GFP</i> background	this study

5.1.10 Media and supplements

Lennox Broth (LB) liquid medium:

- 1% (w/v) tryptone
- 0.5% (w/v) yeast extract
- 85.6 mM NaCl

SPI-2 MES liquid medium (pH 5.8):

- 1x MES buffer
- 1x phosphate buffer
- 0.4% (w/v) glucose
- 15 mM NH₄Cl
- 1 mM MgSO₄
- 10 μ M CaCl₂
- 0.04% (w/v) L-histidine
- 0.001% (w/v) thiamine
- 10x Micronutrients

SOCS liquid medium:

- 2% (w/v) tryptone
- 0.5% yeast extract
- 85.6 mM NaCl
- 2.5 mM KCl
- 10 mM MgCl₂
- 20 mM glucose

Lennox agar:

- LB medium (see above)
- 1.2% (w/v) agar

TOP agar:

- 1% (w/v) tryptone
- 1% (w/v) agar
- 10 mM MgSO₄
- 5 mM CaCl₂
- 86 mM NaCl

DMEM complete:	DMEM (Gibco) 10% fetal calf serum (FCS; Biochrom) 2 mM L-glutamine (Gibco) 1 mM sodium pyruvate (Gibco)
DMEM complete + Pen/Strep:	DMEM complete 1% penicillin/streptomycin
DMEM complete + gentamicin (low):	DMEM complete 10 µg/mL gentamicin
DMEM complete + gentamicin (high):	DMEM complete 50 µg/mL gentamicin
RPMI complete:	RPMI (Gibco) 10% fetal calf serum (FCS; Biochrom) 2 mM L-glutamine (Gibco) 1 mM sodium pyruvate (Gibco) 0.5% β-mercaptoethanol
RPMI complete + Pen/Strep:	RPMI complete 1% penicillin/streptomycin
RPMI complete + gentamicin (low):	RPMI complete 10 µg/mL gentamicin
RPMI complete + gentamicin (high):	RPMI complete 50 µg/mL gentamicin
X-vivo-15 complete:	X-vivo-15 medium (Lonza) 10% FCS 10% L929-conditioned DMEM (composition as above)
X-vivo-15 + gentamicin (low):	X-vivo-15 complete 10 µg/mL gentamicin
X-vivo-15 + gentamicin (high):	X-vivo-15 complete 50 g/mL gentamicin

5.1.11 Buffers and solutions

30:1 ethanol/sodium acetate: (for RNA precipitation)	30 parts of 100% ethanol 1 part of 3M sodium acetate (pH 6.5)
---	--

Agarose gel electrophoresis solution:	X% (w/v) agarose in 1x TAE buffer
Chemiluminescence solution A:	0.1 M Tris-HCl (pH 8.6) 0.025% (w/v) luminol
Chemiluminescence solution B:	0.11% (w/v) <i>p</i> -coumaric acid (in DMSO)
DNA loading dye (5x stock):	10 mM Tris-HCl (pH 7.6) 60% (v/v) glycerol 60 mM EDTA (pH 8.0) 0.025% (w/v) bromphenol blue
PAA gel electrophoresis solution (6%): (for Northern blots)	100 mL 10x TBE 420 g 7M Urea 150 mL Rotiphorese Gel 40 (19:1) H ₂ O ad 1 L
PAA stacking gel solution (4%): (for Western blots)	1.25 mL Tris solution ("upper buffer") 1 mL Rotiphorese Gel 40 (37.5:1) 7.5 mL H ₂ O 75 µL 10% (w/v) APS 7.5 µL TEMED
PAA separation gel solution (12%): (for Western blots)	3.75 mL Tris solution ("lower buffer") 3 mL Rotiphorese Gel 40 (37.5:1) 3.25 mL H ₂ O 150 µL 10% (w/v) APS 15 µL TEMED
Protein loading dye RPA (2x stock):	98% (v/v) formamid 2 mM EDTA (pH 8.0) 0.02% (w/v) bromphenol blue 0.02% (w/v) xylene cyanol
RNA loading dye GL-II (2x stock):	0.13% (w/v) SDS 18 µM EDTA (pH 8.0) 95% formamid 0.025% (w/v) bromphenol blue 0.025% (w/v) xylene cyanol
SDS running buffer (10x stock):	30.275 g Tris base 144 g glycine 10 g SDS H ₂ O ad 1 L

SSC buffer (20x stock):	3 M NaCl 0.3 M sodium citrate titrate to pH 7.0 (using HCl)
TAE buffer (50x stock):	242 g Tris base 51.7 mL acetic acid 10 mM EDTA (pH 8.0) H ₂ O ad 1 L
TBE buffer (10x stock):	108 g Tris base 55 g boric acid 20 mM EDTA (pH 8.0) H ₂ O ad 1 L
TBS buffer (10x stock):	24.11 g Tris base 72.6 g NaCl titrate to pH 7.4 (using HCl) H ₂ O ad 1 L
TBST buffer (10x stock):	1x TBS 0.1% (v/v) Tween-20
Transfer buffer (10x stock):	30 g Tris base 144 g glycine H ₂ O ad 1 L
Tris “lower buffer” solution:	1.5 M Tris-HCl (pH 8.8) 0.4% (w/v) SDS
Tris “upper buffer” solution:	0.5 M Tris-HCl (pH 6.8) 0.4% (w/v) SDS
Western development solution:	2 mL chemiluminescence solution A 200 chemiluminescence solution B 5 µL 3% (v/v) H ₂ O ₂

5.2 Methods

5.2.1 Microbiological methods

Standard growth conditions: *Salmonella enterica* serovar Typhimurium strain SL1344 constitutively expressing GFP from a chromosomal locus (strain JVS-3858) was previously described (Papenfors et al., 2009) and is referred to as wild-type throughout this study. The

complete list of bacterial strains used in this study is provided in Table 5.9. Bacteria were grown in Lennox broth (LB) medium at 37°C with shaking at 220 rpm. When appropriate, 100 µg/mL ampicillin (Amp), 50 µg/mL kanamycin (Kan) or 20 µg/mL chloramphenicol (Cm) (final concentrations) were added to the liquid medium or agar plates.

Growth under SPI-2 conditions: To grow *Salmonella* under SPI-2-inducing conditions, 1 mL of a LB overnight cultures was washed 2x in PBS and 1x in synthetic SPI-2 medium (Lober et al., 2006). The resulting pellet was resuspended in 1 mL of SPI-2 medium and diluted 1:50 in fresh, pre-warmed SPI-2 medium (10 mL total culture volume) in 100 mL Erlenmeyer flasks. The culture was grown at 37°C, 220 rpm until it reached an OD₆₀₀ of 0.3 (takes approximately 3 h).

In vitro SPI-1 to SPI-2 switch assay: To mimic the early stages of the infection of a host cell *in vitro*, the indicated *Salmonella* strains were grown in LB overnight, diluted 1:100 in LB and grown to an OD₆₀₀ of 2 (i.e. a condition under which SPI-1 is highly induced (Kroger et al., 2013; Sittka et al., 2007)), washed 2x with PBS and 1x with SPI-2 medium, diluted 1:50 in pre-warmed SPI-2 medium and transferred to either an Erlenmeyer flask (10 mL culture volume; for Western blotting) or each three wells of a 96-well plate (150 µL culture volume; for measuring GFP intensity). In SPI-2 medium the bacteria were kept growing for the indicated time periods. At the respective time points total protein samples were taken for Western blotting or GFP fluorescence was measured.

Transformation of chemically competent *E. coli*: 10 µL of chemically competent *E. coli* TOP10 cells (Invitrogen) were mixed with 2 µL of a ligation reaction or 10-50 ng of plasmid DNA. The mixture was pre-incubated for 30 min on ice. Subsequently cells were heat-shocked for 30 s at 42°C and chilled for 1 min on ice. Then 900 µL of LB medium was added and cells were recovered for 1 h at 37°C, 220 rpm. 100 µL of the culture and the residual volume were plated on LB agar supplemented with the appropriate selection antibiotics.

Preparation and transformation of electro-competent *Salmonella*: In order to prepare electro-competent *Salmonella* cells, LB cultures were inoculated with bacteria from an overnight culture (1:100 dilution) and grown at 37°C, 220 rpm to an OD₆₀₀ of 0.5. Subsequently, the cells were chilled for 30 min on ice and centrifuged (20 min; 4,000 rpm; 4°C). The cell pellets were next washed once with ice-cold H₂O and twice with 10% (v/v) glycerol. The pellets were resuspended in 1 OD₆₀₀/100 µL of ice-cold H₂O and each 100 µL were distributed in pre-cooled electroporation cuvettes (2 mm gap size) and mixed with ~10 ng of the respective plasmid DNA. Transformation was achieved by electroporation (200 Ω; 25 µF; 2.5 kV). Transformed bacteria were resuspended in 900 µL of SOC medium and recovered by shaking at 220 rpm and 37°C for 1 h.

One-step modification of genes in the chromosome: Chromosomal mutagenesis of *Salmonella* SL1344 was performed as previously described (Datsenko and Wanner, 2000). To construct the STnc440 deletion strain (JVS-10038), the first ~60 nt of the gene were removed while keeping the Rho-independent terminator intact. The respective resistance cassette was eliminated using the FLP helper plasmid pCP20 at 42°C (Datsenko and Wanner, 2000). All mutations were transduced into the wild-type background using phage P22 (Sternberg and Maurer, 1991).

Phage P22 transduction: For the transfer of chromosomal mutations to different strain backgrounds, P22 phage lysates were prepared using soft agar plates as previously described (Sternberg and Maurer, 1991). Briefly, 100 µl of the culture of the respective donor strain (OD₆₀₀ of 1.0) were added to 3 mL of TOP agar containing 20 mM MgSO₄, 10 mM CaCl₂ and the mixture was poured onto a pre-warmed LB agar plate. 100 µL of the P22 lysate from wild-type bacteria were plated on top of the TOP agar. Plates were incubated at 37°C overnight. The next day, the TOP agar layer was scraped off, transferred to a glass tube and resuspended in 5 mL of liquid LB/10 mM MgSO₄/5 mM CaCl₂ and supplemented with 400 µL of chloroform. The resulting suspension was mixed thoroughly, incubated overnight at 4°C, and subsequently centrifuged for 10 min at 4,000 rpm (4°C). The resulting supernatant fraction was transferred to a fresh glass tube containing 400 µL of chloroform for storage at 4°C.

For transduction, 100 µL of a culture of the recipient strain (OD₆₀₀ of 1.0) were mixed with 1-50 µL of the phage lysate (see above) and incubated for 20 min at room temperature. Addition of EGTA to a final concentration of 10 mM was used to stop the transduction. The sample was then plated on pre-warmed LB agar plates supplemented with appropriate antibiotics to select the positive transductants, which were subsequently verified by PCR.

sRNA pulse-expression (in vitro and in vivo): For the *in vitro* experiment in SPI-2 medium, arabinose-induced over-expression of STnc440 from the pBAD plasmid was achieved using the previously described protocol (Papenfort et al., 2006; Papenfort et al., 2012; Papenfort et al., 2009) with minor modifications. Briefly, wild-type *Salmonella* that had been transformed with pKP8-35 (pBAD control), pYC5-34 (pBAD-STnc440) or pYC60 (pBAD-STnc440*) were grown overnight in LB. The next day, the cultures were washed 2x with PBS and 1x with SPI-2 medium, diluted 1:50 in SPI-2 medium and grown to an OD₆₀₀ of 0.3. Then L-arabinose (Sigma) was added to the culture in a final concentration of 0.2%. 5 min after induction, total RNA was extracted using the TRIzol LS reagent (Invitrogen). cDNA libraries were prepared as explained below and analyzed by RNA-seq on Illumina's MiSeq machine (~3-5 M reads/library).

For the pulse-expression of STnc440 inside host cells, HeLa-S3 cells were infected (m.o.i. of 5) with the same three strains and 4 h after infection, 0.2% L-arabinose was supplemented into the DMEM medium. Activation of inducible sRNA expression in intracellular bacteria was confirmed

by qRT-PCR over a time-course of 20 min, demonstrating full induction levels to be reached at 5 min. Thus, 5 min after induction the host cells were lysed with ice-cold 1% Triton X-100/PBS and further incubated for 30 min on ice with pipetting up and down from time to time to improve lysis efficiency. Then the intact bacterial cells were pelleted by centrifugation for 2 min at 16,100 g (4°C) and resuspended in *RNAlater* (Qiagen). The thus fixed bacterial cells were further enriched against the host background via cell-sorting (FACS Aria III, BD Biosciences) and selective gating for the fraction of GFP⁺ bacterial cells released from their hosts. Total RNA was isolated and analyzed by RNA-seq as above except that sequencing was on the HiSeq platform to ~20 M reads/library as despite the enrichment of bacterial cells the majority of sequence reads were still host-derived.

EGFP fusion assays (flow cytometry, plate reader): For sRNA target validation, the CDS of EGFP was fused downstream of and in frame with the respective target sequence (5' UTR plus early coding sequence) and inserted into plasmid pXG-10. *Salmonella* transformed with either the empty plasmid or the target gene reporter construct in combination with the pBAD-ctrl., pBAD-STnc440 or pBAD-STnc440* were grown in LB (Amp + Cm) overnight. The next day, 50 µL of the culture were diluted into 2 mL LB (Amp + Cm + 0.2% L-arabinose) and grown for 3 h to ~OD₆₀₀ of 2. Then, 1 OD₆₀₀ bacteria were collected for measurement of GFP fluorescence (FACSCalibur, BD Biosciences).

To measure SPI-2 activation, a transcriptional fusion of the *ssaG* promoter to the CDS of EGFP was inserted into plasmid pAS046 via *AatII*/*NheI* sites as previously described (Pfeiffer et al., 2007). The resulting reporter construct or the empty control plasmid pXG1 were co-transformed with expression plasmids pBAD-ctrl., pBAD-STnc440 or pBAD-STnc440* into the indicated strain backgrounds. The resulting strains were grown overnight in LB (Amp + Cm) and then diluted 1:100 and further in LB to an OD₆₀₀ of 2. Each 1 mL of the cultures was pelleted and cells shifted to SPI-2 medium (defined as t₀) as described above, except that the growth experiment was conducted in 96-well plates (TC Microwell 96F, Thermo Scientific). After measuring the OD₆₀₀ and GFP intensity at t₀, L-arabinose was added to each well to final concentration of 0.2% and bacteria were grown for 20 h at 37°C (shaking) with measurements of both the OD₆₀₀ and GFP fluorescence in 10 min intervals using the Infinite F200 PRO plate reader (TECAN). For each technical triplicate the ratios of GFP intensity and OD₆₀₀-value were calculated, normalized against strains carrying the empty vector control (pXG1) and are expressed relative to the maximum reached in the wild-type background.

5.2.2 Cell culture methods

Passaging and seeding: The following cell lines were used in this study: Human cervix carcinoma (HeLa-S3; ATCC CCL-2.2), human epithelial colorectal adenocarcinoma cells (CaCo-2; ATCC HTB-

37), human stomach adenocarcinoma cells (AGS; ATCC CRL-1739), human embryonic kidney 293 cells (HEK 293; ATCC CRL-1573), human monocytic cells (THP-1; ATCC TIB-202), murine embryonic fibroblast cells (MEF; ATCC SCRC-1040), mouse leukaemic monocyte/macrophage cell line (RAW264.7; ATCC TIB-71). All cell lines except for AGS cells were obtained from the group of Thomas Rudel (Biocenter Würzburg). AGS cells were provided by Cynthia Sharma (ZINF Würzburg). HeLa-S3 cells were cultured according the guidelines provided by the ENCODE consortium (genome.ucsc.edu/encode/protocols/cell/human/Stam_15_protocols.pdf). Briefly, cells were grown in DMEM (Gibco) supplemented with 10% fetal calf serum (FCS; Biochrom), 2 mM L-glutamine (Gibco) and 1 mM sodium pyruvate (Gibco) in T-75 flasks (Corning) in a 5% CO₂, humidified atmosphere at 37°C. All further cell lines used in this study (THP-1, CaCo-2, AGS, HEK 293, MEF and RAW264.7) were cultured in RPMI (Gibco) supplemented with 10% FCS, 2 mM L-glutamine, 1 mM sodium pyruvate in a 5% CO₂, humidified atmosphere at 37°C. To differentiate THP-1 monocytes, seeded cells (1x10⁶ cells/well; six-well format) were treated with 50 ng/mL (final concentration) of phorbol 12-myristate 13-acetate (PMA) (Sigma) for 72 h (after 48 h fresh PMA in the same concentration was added to the culture). For the differentiation of bone marrow derived macrophages (BMDMs) the marrow of femur and tibia was isolated from 8-12 week old female C57BL/6 wild-type mice and stored in RPMI supplemented with 10% FCS. The cell suspension was centrifuged for 5 min at 250 g and the leukocyte pellet was resuspended in differentiation medium consisting of X-vivo-15 medium (Lonza) supplemented with 10% FCS and 10% L929 conditioned DMEM medium (same composition as above). Cells were cultured for 3 days at 3x10⁶ cells per 10 mL in a T-75 flask. At day 3 another 3 mL of differentiation medium were added and cells were further cultured until day 5. Successful macrophage differentiation was validated by microscopy on day 5 upon which the cells were detached using a rubber scraper (Sarstedt) and seeded into six-well plates at 10⁵ cells per well in fresh differentiation medium. Infection was carried out on day 7 as described below.

Apoptosis assay: To detect apoptotic cells, HeLa-S3 were washed twice with PBS and resuspended in 1x binding buffer (BD Pharmingen) in a concentration of 10⁶ cells/mL. 100 µL of this cell suspension were then mixed with 5 µL of APC-labeled Annexin V (BD Pharmingen) and 1 µL of 500 mg/mL propidium iodide (PI; lyophilized stock from Sigma). Upon incubation for 15 min at room temperature (light-protected) cells were subjected to flow cytometry (see below).

LDH release assay: Necrosis was evaluated by quantifying released lactate dehydrogenase (LDH) via the Cytotox96 assay (Promega) according to the manufacturer's instructions. The absorbance at 490 nm was measured using a Multiskan Ascent instrument (Thermo Fisher). In order to convert the measured absorbance values into the relative proportion of dead cells, the maximal absorbance was determined by using 1x Lysis Solution (Promega) following the

manufacturer's instructions and referred to as 100% cytotoxicity. For both apoptosis and cytotoxicity measurements at least three biological replicates were analyzed, each of which comprised three technical replicates.

5.2.3 Infection assays

In vitro Salmonella infection assay: Infection of HeLa-S3 cells was carried out following the protocol of (Schulte et al., 2011) with slight modifications. Two days prior to infection 2×10^5 HeLa-S3 cells were seeded in 2 mL complete DMEM (six-well format). Overnight cultures of *Salmonella* were diluted 1:100 in fresh LB medium and grown aerobically to an OD₆₀₀ of 2. Bacterial cells were harvested by centrifugation (2 min at 12,000 rpm, room temperature) and resuspended in complete DMEM medium. Infection of HeLa-S3 cells was carried out by adding the bacterial suspension directly to each well. If not mentioned otherwise, infections were performed at a multiplicity of infection (m.o.i.) of 5. Immediately after addition of bacteria, the plates were centrifuged for 10 min at 250 g, room temperature followed by 30 min incubation in 5% CO₂, humidified atmosphere, at 37°C. Medium was then replaced for gentamicin-containing complete DMEM (final concentration: 50 mg/mL) to kill extracellular bacteria. After a further 30 min incubation step, medium was again replaced by 10 mg/mL gentamicin/DMEM and incubated for the remainder of the experiment. Time point 0 h was defined as the time when gentamicin was first added to the cells.

Further cell types (THP-1, CaCo-2, AGS, HEK 293, MEF and RAW264.7) were infected as described for HeLa-S3 cells except that infection was carried out in RPMI complete medium instead of DMEM and that infection was with a m.o.i. of 10 instead of m.o.i. 5. Infection of BMDMs was carried with a m.o.i. of 10 and using X-vivo-15 medium (10% fetal calf serum, 10% L929 conditioned medium).

Quantification of intracellular replication: Infected HeLa-S3 cells were analyzed by flow cytometry as described below. Alternatively, infected HeLa-S3 cultures were solubilized with PBS containing 0.1% Triton X-100 (Gibco) at the respective time points. Cell lysates were serially diluted in PBS, plated onto LB plates and incubated at 37°C overnight. The number of colony forming units (c.f.u.) recovered was compared to that obtained from the bacterial input solution used for infection. In all cases, at least three biological replicates were analyzed, each of which comprised three technical replicates.

5.2.4 Basic molecular biological methods

Preparation of plasmid DNA: Plasmid DNA was extracted from transformed TOP10 cells using the NucleoSpin Plasmid QuickPure kit (Macherey-Nagel) following the manufacturer's instructions.

PCR program:	98°C	2 min (for colony-PCR: 10 min)
	98°C	30 sec
	55-60°C	30 sec
	72°C	30 sec/1 kb
	72°C	5 min
	15°C	∞

Resulting PCR products were either purified directly using the QIAquick PCR Purification kit (Qiagen) following the manufacturer's instructions or separated by AGE, excised from the gel and purified using the QIAquick Gel Extraction Kit (Qiagen).

Restriction digestion, DpnI digestion and DNA ligation: Restriction digestions were performed using restriction enzymes and buffers from NEB and Fermentas according to the instructions given by the manufacturers. Plasmid products from mutagenesis PCR were treated for 1 h at 37°C with DpnI (NEB) to remove remnants of the template plasmid. Blunt and sticky end ligation of DNA fragments was achieved using the T4 DNA ligase (NEB) according to the manufacturer's recommendations.

Agarose gel electrophoresis (AGE): After PCR or restriction digestion, individual DNA fragments were separated by AGE. Gels were prepared by dissolving pure agarose in 1×TAE buffer to the desired concentration (typically 1-2% (w/v)). The solution was boiled until all agarose was dissolved, supplemented with 0.02% (v/v) of the RedSafe dye (ABC Scientific) and casted. Samples were supplemented with ¼ volume of 5x DNA loading buffer, mixed and loaded on the gel. The GeneRuler 1 kb DNA Ladder (Fermentas) was loaded as a size marker. Separation was at 100-150 V.

Denaturing polyacrylamide gel electrophoresis (PAGE): To size-separate individual RNA species, denaturing polyacrylamide (PAA) gels were used. Prior to casting the gels, all glass ware and equipment was cleaned with 70% ethanol. The 40% PAA solution (19:1 acrylamide/bis-acrylamide) was diluted in water to the desired concentration and supplemented with urea to a final concentration of 8.3M. The addition of 0.01 volumes of 10% APS and 0.001 volumes of TEMED polymerized the PAA solution. RNA samples were supplemented with 1 volume of 2x GL-II loading dye, denatured by boiling for 5 min at 100°C and chilled on ice for 5 min. 1x TBE was used as a running buffer. PAGE was performed at 300 V and room temperature.

5.2.5 RNA techniques

Total RNA extraction: For Dual RNA-seq, RNA was extracted from cells using the *mirVana* kit (Ambion) following the manufacturer's instructions for total RNA isolation. This method had been optimized to recover transcripts of all sizes, including short RNA fractions (in particular miRNAs and sRNAs) in addition to larger transcripts, and in this respect may surpass traditional column-based RNA isolation techniques (Stead et al., 2012). To this end, infected cells at a given time point were trypsinized, collected and directly lysed in 600 µL of Lysis/Binding buffer of the *mirVana* kit (Ambion). As a reference for gene expression changes in HeLa-S3 upon infection, a non-infected yet mock-treated control was included. The bacterial reference samples were derived from *Salmonella* grown in LB to an OD₆₀₀ of 2, which either were then shifted to

complete DMEM for 15 min, pelleted and fixed in RNAlater (see below) or were fixed directly (i.e. without a medium exchange step). Fixed *Salmonella* cells were pelleted and lysed in Lysis/Binding buffer. In order to maintain the approximate ratio of bacterial to human transcripts during RNA isolation, *Salmonella* lysates were mixed with HeLa-S3 lysate in a way that the calculated proportion of individual *Salmonella* cells per infected host cell at the respective time points was matched. The resulting mixture was then processed collectively.

For qRT-PCR or Northern blot experiments as well as for RNA-seq-based analysis of pulse-expression experiments, total RNA was extracted using the TRIzol LS reagent (Invitrogen) according to the manufacturer's recommendations.

DNase I treatment: To get rid of contaminating genomic DNA, RNA samples supposed to be analyzed by RNA-seq or qRT-PCR were treated with 1 U of DNase I (Fermentas) per 4 µg of RNA for 45 min at 37°C. If applicable, RNA quality was checked on the Agilent 2100 Bioanalyzer (Agilent Technologies).

qRT-PCR analysis: qRT-PCR experiments were performed with the Power SYBR Green RNA-to-CT 1-Step kit (Applied Biosystems) according to the manufacturer's instructions. Fold-changes were determined using the $2^{-\Delta\Delta Ct}$ -method (Livak and Schmittgen, 2001). Primer sequences are given in Table 5.7 and their specificity had been confirmed using Primer-BLAST (NCBI). For the estimation of *Salmonella* RNA within infection samples (Fig. 2.8), a dilution series of separately isolated *Salmonella* and HeLa-S3 total RNA was set up and in each case the ratio of *rfaH*/ β -actin mRNAs was determined. The same was done for biological samples from infected cells (with or without sorting) as well as for the *Salmonella* reference controls (see above). From the resulting trend-line equation the approximate proportion of the *Salmonella* transcriptome within mixed pro-/eukaryotic total RNA samples could be deduced.

Northern blotting: Total RNA was prepared using the TRIzol LS reagent and separated in denaturing PAA gels as described above. 5-10 µg of RNA per lane were loaded for bacteria only samples and 2 µg (sorted samples) to 50 µg (unsorted) per lane for mixed human/*Salmonella* RNA samples. Northern blotting was as described (Sittka et al., 2007). For detection, Hybond XL membranes (Amersham) were hybridized at 42°C with gene-specific [³²P] end-labeled DNA oligonucleotides (see Table 5.7 for sequences) in Rapid-hyb buffer (GE Healthcare).

5.2.6 Protein techniques

Western blotting: Immunoblotting of *Salmonella* proteins was done as previously described (Urban and Vogel, 2007). Briefly, samples from *Salmonella in vitro* cultures were taken according to 0.4 OD₆₀₀, centrifuged for 4 min at 16,100 g at 4°C, and pellets resuspended in protein loading buffer to a final concentration of 0.01 OD/µL. After denaturation for 5 min at 95°C, 0.05-OD

equivalents of the sample were separated via SDS-PAGE. Gel-fractionated proteins were blotted for 90 min (0.2 mA/cm²; 4°C) in a semi-dry blotter (Peglab) onto a PVDF membrane (PerkinElmer) in transfer buffer (25 mM Tris base, 190 mM glycine, 20% methanol). Blocking was for 1 h at room temperature in 10% dry milk/TBST. Appropriate primary antibodies were hybridized at 4°C overnight and – following 3x 5 min washing in TBST – secondary antibodies for 1 h at room temperature. Primary and secondary antibodies used and the respective dilution are listed in Table 5.6. For Western blotting of human proteins, infected cells were harvested in protein loading buffer (500 µL/well; six-well format), transferred to 1.5 mL reaction tubes, boiled for 5 min at 95°C and 20 µL/lane were loaded onto a 10% PAA gel for SDS-PAGE as above. After blotting and blocking, the membrane was probed with the respective primary antibody at 4°C overnight and for 1 h at room temperature with the corresponding secondary antibodies. Blots were developed using Western Lightning solution (PerkinElmer) in a Fuji LAS-4000. Where indicated, intensities of bacterial protein bands were quantified using the AIDA software (Raytest, Germany) and normalized to GroEL levels.

ELISA: HeLa-S3 cells were infected with wild-type *Salmonella*, Δ STnc440 or Δ STnc440+ mutant strains at a m.o.i. of 5 as described above. Culture supernatant samples were taken at 20 h p.i. and analyzed using the ELISA kit for human CXCL8/IL-8 (R&D Systems).

5.2.7 Fixation techniques

Infected cells were washed twice with PBS, trypsinized and pelleted. For ethanol fixations, cell pellets were re-dissolved in 0.1 volumes of ice-cold PBS and then 0.9 volumes of ice-cold ethanol (either 70% or 100%; as indicated) were added in single droplets during shaking (400 rpm, 4°C) to avoid cell clumping. Fixation using the stop solution (95% EtOH/5% water-saturated phenol; (Eriksson et al., 2003)) was performed by resuspending the cell pellet in PBS prior to the addition of 0.2 volumes of stop solution and mixing. When PFA was used, the pellet was resuspended in the respective PFA concentration (0.5% or 4% (w/v) PFA, pH 7.4, with or without 4% (w/v) sucrose) and shaken for 15 min at 400 rpm, room temperature. PFA-induced cross-links were reverted by an additional heating step for 15 min at 70°C (Hamatani et al., 2006; Kuramochi et al., 2006). For fixation with *RNAlater* (Qiagen) cell pellets were directly resuspended in *RNAlater* (1 mL/5x10⁶ cells) and incubated for 30 min at room temperature. For systematic evaluation of different fixation protocols (Fig. 2.5 and 2.6), fixed cells had not been sorted but were either directly analyzed upon fixation (30 min) or stored at -20°C (ethanol-based fixatives) or 4°C (others), respectively, overnight. *RNAlater*-fixed samples that were to be analyzed by Dual RNA-seq were sorted. To this end, tubes containing ~5x10⁶ fixed cells were filled up with 10 mL of ice-cold PBS, centrifuged (5 min, 500 g, 4°C) and cell pellets resuspended in 1 mL of cold PBS. This cell suspension was filtered and sorted.

5.2.8 Flow cytometry and FACS-based methods

Flow cytometry-based analyses: To measure the infection rate, infected HeLa-S3 cultures were washed twice with PBS, detached from the bottom of the plate by trypsinization and resuspended in complete DMEM. Upon pelleting the cells (5 min at 250 g, room temperature) they were resuspended in PBS and analyzed by flow cytometry using a FACSCalibur instrument (BD Biosciences) and the Cyflogic software (CyFlo Ltd.). Selection of intact HeLa-S3 cells was achieved by gating based on cell diameter (forward-scatter) and granularity (side-scatter) (linear scale). Of those, infected (GFP-positive) and non-infected (GFP-negative) sub-fractions were defined based on GFP signal intensity (FITC channel) vs. auto-fluorescence (PE channel) (logarithmic scale).

To quantify the relative amount of intracellular pathogens per host cell over time, the increase in GFP intensity (geometric mean) was measured in the GFP-positive sub-population and normalized to that of the non-infected population from the same sample as described above.

To measure apoptosis, Annexin V-stained cells were analyzed by flow cytometry using the MACSQuant Analyzer (Miltenyi Biotec). Upon gating of the fraction of intact cells based on cell diameter (forward-scatter) and granularity (side-scatter), the Annexin-positive/PI-negative sub-population was determined by comparison against the appropriate single-stained controls in the APC vs. PerCP channels and quantified.

Cell sorting: RNAlater-fixed cells (see below) were first passed through MACS Pre-Separation Filters (30 μ m exclusion size; Miltenyi Biotec.) and then analyzed and sorted using the FACSARIA III device (BD Biosciences) under continuous cooling to 4°C (both the input tube holder and the collection tube rack) and at a medium flow rate (\sim 4) using the same gating strategy as described above, except that the gates for GFP-positive and -negative fractions were more conservative in order to prevent cross contamination (as exemplified in Fig. 2.7a). Typically $\sim 3 \times 10^5$ cells were collected in each fraction and subjected to RNA isolation.

5.2.9 RNA-seq techniques

Total RNA samples for Dual RNA-seq and pulse-expression were converted into cDNA libraries for Illumina sequencing by Vertis Biotechnologie AG, Freising-Weihenstephan, Germany (<http://www.vertisbiotech.com>). A minimal amount of ~ 100 ng of total RNA was required for cDNA library preparation. DNase I-treated total RNA samples were first sheared via ultra-sound sonication (4 pulses of 30 s at 4°C each) to generate on average ~ 200 -400 bp fragmentation products. Then, fragments < 20 nt were removed using the Agencourt RNAClean XP kit (Beckman Coulter Genomics). As an internal quality control for the pilot experiment, spike-in RNA (5'-AAAUCCGUUCGUACGGCCCC-3'; 5'-monophosphorylated and gel-purified) was added to a final

concentration of 0.5%. The samples were then poly(A)-tailed using poly(A) polymerase and the 5' triphosphate (and eukaryotic cap) structures were removed using tobacco acid pyrophosphatase (TAP). Afterwards, a RNA adapter was ligated to the 5' monophosphate of the RNA. First-strand cDNA synthesis was performed using an oligo(dT)-adapter primer and the M-MLV reverse transcriptase (NEB). The resulting cDNA was PCR-amplified to about 20-30 ng/ μ L using a high fidelity DNA polymerase (barcode sequences for multiplexing were part of the 3' primers). The cDNA library was purified using the Agencourt AMPure XP kit (Beckman Coulter Genomics) and analyzed by capillary electrophoresis (Shimadzu). Single-end sequencing (100 cycles) was performed on an Illumina HiSeq 2000 machine at the Max Planck Genome Centre Cologne, Cologne, Germany or on the Illumina MiSeq platform at the Institute for Molecular Infection Biology, University of Würzburg, Würzburg, Germany.

5.2.10 Microscopy techniques

For microscopic analyses, infection was carried out as described above, except that HeLa-S3 cells were seeded onto coverslips (24-well format). At the respective time point coverslips with infected HeLa-S3 were washed twice with PBS (Gibco) and fixed in 4% (w/v) PFA for 15 min in a wet chamber. After two additional PBS washing steps, cells were stained with Hoechst 33342 (Invitrogen; 1:5,000 diluted in PBS) for 15 min in a wet chamber and again washed twice with PBS. After coverslips had been air-dried, they were embedded in VECTASHIELD(R) Mounting Medium (Biozol) and analyzed using the LEICA SP5 confocal microscope (Leica) and the LAS AF Lite software (Leica).

5.2.11 Bioinformatics analyses

RNA-seq data processing: Illumina reads in FASTQ format were trimmed with a phred score cut-off of 20 by the program fastq_quality_trimmer from the FASTX toolkit version 0.0.13 (http://hannonlab.cshl.edu/fastx_toolkit/). For the high-resolution time-course experiment (as outlined in Figure 3.2), the FASTQ files were sub-sampled to have an equal number of 146 M total reads/library except for the 0 h bacterial reference samples which were sub-sampled to 10 M reads mapping to the *Salmonella* genome. To compare the Δ STnc440 data to that of other sRNA mutants (Fig. 4.3 and 4.4), the respective time points from the deeply sequenced experiment were sub-sampled to 30 M total reads to ensure an equal sequencing depth for all libraries that were compared. For read mapping, 5' linker-clipped reads were aligned to the *Salmonella enterica* SL1344 genome (accession numbers: NC_016810.1, NC_017718.1, NC_017719.1, NC_017720.1) and the human (hg19 – hs37d5 retrieved from the 1000 Genome Project (Abecasis et al., 2012)) or the mouse genome sequence (GENCODE M2), respectively, in parallel using the READemption pipeline (Förstner et al., 2014) and the short read mapper segemehl (Otto et al., 2014) allowing for split reads (Hoffmann et al., 2014). Reads shorter than

20 nt after poly(A)-trimming were discarded prior to the mapping. Mapped reads with an alignment accuracy <90% as well as cross-mapped reads – i.e. reads which could be aligned equally well to both human and *Salmonella* reference sequences – were discarded. Based on the resulting BAM files, coverage files in wiggle format were generated by READemption and split by organism. Gene-wise quantifications for the human data subset were performed based on annotations from GENCODE (Harrow et al., 2012), NONCODE (Xie et al., 2014) and miRBase (Kozomara and Griffiths-Jones, 2014) after removing redundant entries and the annotation for *Salmonella* genes was retrieved from NCBI (under the above mentioned accession numbers) and manually extended by small RNA annotations (Förstner et al., unpublished). Based on gene quantifications provided by the READemption pipeline the calculation of RPKM values and all subsequent analysis steps were carried out separately for the host and the pathogen system. Different RPKM cut-off values did not alter the general trends observed. All differential expression analysis was performed using the DESeq R package (Anders and Huber, 2010) with method blind and sharing mode fit-only.

Salmonella analyses: A database of pathways, regulons, and genomic islands was constructed using information obtained from the KEGG database (Kanehisa et al., 2014) (organism code sey), the SL1344 genome annotation (Kroger et al., 2012), and relevant literature sources. Spearman correlation coefficients between changes in STnc440 expression and changes in expression of each gene within each regulon over the time-course were plotted in Figure 3.6. To assess enrichment of differentially expressed transcripts in pathways (Figure 3.9), genes displaying an absolute fold-change of at least 2 and a RPKM value of at least 20 in both the wild-type and Δ STnc440 were considered differentially expressed, and a hypergeometric test for enrichment was conducted for each pathway. To generate Figure 4.3a, aligned reads for the STnc440 deletion and control libraries were subsampled to be comparable in depth to those of the other sRNA deletion libraries. Transcripts with a minimum count per million reads excluding ribosomal and tRNAs of at least 400 (approximately 20 reads) in either the wild type or deletion strain and a fold-change of at least 2 in at least one deletion:wild-type comparison are displayed.

Human analyses: A transcript was considered to be detected if in at least one condition the observed RPKM was above 0.1 for lncRNA genes and 0.5 for all other human transcript classes, respectively. For a given comparison, all detected transcripts showing at least a 1.5 fold-change in abundance and a p-value below 0.01 were considered differentially expressed.

Statistics: Tests for the evaluation of statistical significance in Figure 2.3 were performed using a one-tailed Student's t-test. p-values <0.05 were considered statistical significant (*) and p-values <0.001 were considered very significant (***). The significance of differences in qRT-PCR results in Figure 3.10 was assessed using a one-tailed Mann-Whitney U test.

6. References

- Abdullah, Z., and Knolle, P.A. (2014). Scaling of immune responses against intracellular bacterial infection. *EMBO J*.
- Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature* 491, 56-65.
- Albrecht, M., Sharma, C.M., Dittrich, M.T., Muller, T., Reinhardt, R., Vogel, J., and Rudel, T. (2011). The transcriptional landscape of *Chlamydia pneumoniae*. *Genome Biol* 12, R98.
- Albrecht, M., Sharma, C.M., Reinhardt, R., Vogel, J., and Rudel, T. (2010). Deep sequencing-based discovery of the *Chlamydia trachomatis* transcriptome. *Nucleic Acids Res* 38, 868-877.
- Almada, A.E., Wu, X.B., Kriz, A.J., Burge, C.B., and Sharp, P.A. (2013). Promoter directionality is controlled by U1 snRNP and polyadenylation signals. *Nature* 499, 360-U141.
- Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biology* 11.
- Aravin, A.A., Hannon, G.J., and Brennecke, J. (2007). The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science* 318, 761-764.
- Armour, C.D., Castle, J.C., Chen, R., Babak, T., Loerch, P., Jackson, S., Shah, J.K., Dey, J., Rohl, C.A., Johnson, J.M., *et al.* (2009). Digital transcriptome profiling using selective hexamer priming for cDNA synthesis. *Nat Methods* 6, 647-649.
- Asai, S., Rallapalli, G., Piquerez, S.J., Caillaud, M.C., Furzer, O.J., Ishaque, N., Wirthmueller, L., Fabro, G., Shirasu, K., and Jones, J.D. (2014). Expression Profiling during Arabidopsis/Downy Mildew Interaction Reveals a Highly-Expressed Effector That Attenuates Responses to Salicylic Acid. *PLoS Pathog* 10, e1004443.
- Asmann, Y.W., Klee, E.W., Thompson, E.A., Perez, E.A., Middha, S., Oberg, A.L., Therneau, T.M., Smith, D.I., Poland, G.A., Wieben, E.D., *et al.* (2009). 3' tag digital gene expression profiling of human brain and universal reference RNA using Illumina Genome Analyzer. *Bmc Genomics* 10.
- Auweter, S.D., Bhavsar, A.P., de Hoog, C.L., Li, Y., Chan, Y.A., van der Heijden, J., Lowden, M.J., Coombes, B.K., Rogers, L.D., Stoykov, N., *et al.* (2011). Quantitative mass spectrometry catalogues *Salmonella* pathogenicity island-2 effectors and identifies their cognate host binding partners. *J Biol Chem* 286, 24023-24035.
- Babitzke, P., and Romeo, T. (2007). CsrB sRNA family: sequestration of RNA-binding regulatory proteins. *Current Opinion in Microbiology* 10, 156-163.
- Bajaj, V., Lucas, R.L., Hwang, C., and Lee, C.A. (1996). Co-ordinate regulation of *Salmonella typhimurium* invasion genes by environmental and regulatory factors is mediated by control of *hlaA* expression. *Molecular Microbiology* 22, 703-714.
- Balaban, N.Q., Gerdes, K., Lewis, K., and McKinney, J.D. (2013). A problem of persistence: still more questions than answers? *Nature Reviews Microbiology* 11, 587-591.
- Balbontin, R., Fiorini, F., Figueroa-Bossi, N., Casadesus, J., and Bossi, L. (2010). Recognition of heptameric seed sequence underlies multi-target regulation by RybB small RNA in *Salmonella enterica*. *Molecular Microbiology* 78, 380-394.
- Barquist, L., Langridge, G.C., Turner, D.J., Phan, M.D., Turner, A.K., Bateman, A., Parkhill, J., Wain, J., and Gardner, P.P. (2013). A comparison of dense transposon insertion libraries in the *Salmonella* serovars Typhi and Typhimurium. *Nucleic Acids Res* 41, 4549-4564.
- Barthel, M., Hapfelmeier, S., Quintanilla-Martinez, L., Kremer, M., Rohde, M., Hogardt, M., Pfeffer, K., Russmann, H., and Hardt, W.D. (2003). Pretreatment of mice with streptomycin provides a *Salmonella enterica* serovar typhimurium colitis model that allows analysis of both pathogen and host. *Infection and Immunity* 71, 2839-2858.
- Becker, K., Jahrling, N., Saghafi, S., Weiler, R., and Dodt, H.U. (2012). Chemical clearing and dehydration of GFP expressing mouse brains. *PLoS One* 7, e33916.
- Behlau, I., and Miller, S.I. (1993). A Phop-Repressed Gene Promotes *Salmonella*-Typhimurium Invasion of Epithelial-Cells. *J Bacteriol* 175, 4475-4484.

- Beisel, C.L., and Storz, G. (2010). Base pairing small RNAs and their roles in global regulatory networks. *FEMS Microbiol Rev* 34, 866-882.
- Belland, R.J., Zhong, G., Crane, D.D., Hogan, D., Sturdevant, D., Sharma, J., Beatty, W.L., and Caldwell, H.D. (2003). Genomic transcriptional profiling of the developmental cycle of *Chlamydia trachomatis*. *Proc Natl Acad Sci U S A* 100, 8478-8483.
- Benjamin, J.A.M., Desnoyers, G., Morissette, A., Salvail, H., and Masse, E. (2010). Dealing with oxidative stress and iron starvation in microorganisms: an overview. *Can J Physiol Pharm* 88, 264-272.
- Bent, Z.W., Tran-Gyamfi, M.B., Langevin, S.A., Brazel, D.M., Hamblin, R.Y., Branda, S.S., Patel, K.D., Lane, T.W., and VanderNoot, V.A. (2013). Enriching pathogen transcripts from infected samples: A capture-based approach to enhanced host-pathogen RNA sequencing. *Analytical Biochemistry* 438, 90-96.
- Bertone, P., Stolc, V., Royce, T.E., Rozowsky, J.S., Urban, A.E., Zhu, X., Rinn, J.L., Tongprasit, W., Samanta, M., Weissman, S., *et al.* (2004). Global identification of human transcribed sequences with genome tiling arrays. *Science* 306, 2242-2246.
- Bhaya, D., Davison, M., and Barrangou, R. (2011). CRISPR-Cas Systems in Bacteria and Archaea: Versatile Small RNAs for Adaptive Defense and Regulation. *Annual Review Genetics*, Vol 45 45, 273-297.
- Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigo, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., Dermitzakis, E.T., Thurman, R.E., *et al.* (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447, 799-816.
- Bishop, R.E., Leskiw, B.K., Hodges, R.S., Kay, C.M., and Weiner, J.H. (1998). The entericidin locus of *Escherichia coli* and its implications for programmed bacterial cell death. *J Mol Biol* 280, 583-596.
- Bjur, E., Eriksson-Ygberg, S., Aslund, F., and Rhen, M. (2006). Thioredoxin 1 promotes intracellular replication and virulence of *Salmonella enterica* serovar Typhimurium. *Infection and Immunity* 74, 5140-5151.
- Blair, J.M., Richmond, G.E., Bailey, A.M., Ivens, A., and Piddock, L.J. (2013). Choice of bacterial growth medium alters the transcriptome and phenotype of *Salmonella enterica* Serovar Typhimurium. *PLoS One* 8, e63912.
- BlancPotard, A.B., and Groisman, E.A. (1997). The *Salmonella* selC locus contains a pathogenicity island mediating intramacrophage survival. *Embo Journal* 16, 5376-5385.
- Blencowe, B.J., Ahmad, S., and Lee, L.J. (2009). Current-generation high-throughput sequencing: deepening insights into mammalian transcriptomes. *Genes Dev* 23, 1379-1386.
- Bobrovskyy, M., and Vanderpool, C.K. (2014). The small RNA SgrS: roles in metabolism and pathogenesis of enteric bacteria. *Frontiers in cellular and infection microbiology* 4.
- Botsford, J.L., and Harman, J.G. (1992). Cyclic-Amp in Prokaryotes. *Microbiol Rev* 56, 100-122.
- Brumell, J.H., Kujat-Choy, S., Brown, N.F., Vallance, B.A., Knodler, L.A., and Finlay, B.B. (2003). SopD2 is a novel type III secreted effector of *Salmonella typhimurium* that targets late endocytic compartments upon delivery into host cells. *Traffic* 4, 36-48.
- Bruno, V.M., Hannemann, S., Lara-Tejero, M., Flavell, R.A., Kleinstein, S.H., and Galan, J.E. (2009). *Salmonella* Typhimurium type III secretion effectors stimulate innate immune responses in cultured epithelial cells. *PLoS Pathog* 5, e1000538.
- Buchholz, K.R., and Stephens, R.S. (2008). The cytosolic pattern recognition receptor NOD1 induces inflammatory interleukin-8 during *Chlamydia trachomatis* infection. *Infection and Immunity* 76, 3150-3155.
- Bustamante, V.H., Martinez, L.C., Santana, F.J., Knodler, L.A., Steele-Mortimer, O., and Puente, J.L. (2008). HilD-mediated transcriptional cross-talk between SPI-1 and SPI-2. *Proc Natl Acad Sci U S A* 105, 14591-14596.
- Caldelari, I., Chao, Y., Romby, P., and Vogel, J. (2013). RNA-Mediated Regulation in Pathogenic Bacteria. *Cold Spring Harb Perspect Med* 3.
- Caly, D.L., Coulthurst, S.J., An, S.Q., Helaine, S., Malone, J.G., and Ryan, R.P. (2014). Communication, cooperation, and social interactions: a report from the third young

- microbiologists symposium on microbe signalling, organisation, and pathogenesis. *J Bacteriol* 196, 3527-33.
- Carey, A.J., Tan, C.K., Ulett, G.C. (2012). Infection-induced IL-10 and JAK-STAT: A review of the molecular circuitry controlling immune hyperactivity in response to pathogenic microbes. *JAKSTAT* 3, 159-67.
- Carpenter, S., Aiello, D., Atianand, M.K., Ricci, E.P., Gandhi, P., Hall, L.L., Byron, M., Monks, B., Henry-Bezy, M., Lawrence, J.B., *et al.* (2013). A Long Noncoding RNA Mediates Both Activation and Repression of Immune Response Genes. *Science* 341, 789-792.
- Carty, S.M., Sreekumar, K.R., and Raetz, C.R.H. (1999). Effect of cold shock on lipid A biosynthesis in *Escherichia coli* - Induction at 12 degrees C of an acyltransferase specific for palmitoleoyl-acyl carrier protein. *Journal of Biological Chemistry* 274, 9677-9685.
- Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B.M., Strein, C., Davey, N.E., Humphreys, D.T., Preiss, T., Steinmetz, L.M., *et al.* (2012). Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* 149, 1393-1406.
- Cavanagh, A.T., and Wassarman, K.M. (2014). 6S RNA, a Global Regulator of Transcription in *Escherichia coli*, *Bacillus subtilis*, and Beyond. *Annu Rev Microbiol* 68, 45-60.
- Chabelskaya, S., Bordeau, V., and Felden, B. (2014). Dual RNA regulatory control of a *Staphylococcus aureus* virulence factor. *Nucleic Acids Research* 42, 4847-4858.
- Chao, Y., Papenfort, K., Reinhardt, R., Sharma, C.M., and Vogel, J. (2012). An atlas of Hfq-bound transcripts reveals 3' UTRs as a genomic reservoir of regulatory small RNAs. *EMBO J* 31, 4005-4019.
- Chao, Y., and Vogel, J. (2010). The role of Hfq in bacterial pathogens. *Curr Opin Microbiol* 13, 24-33.
- Chaudhuri, R.R., Morgan, E., Peters, S.E., Pleasance, S.J., Hudson, D.L., Davies, H.M., Wang, J.H., van Diemen, P.M., Buckley, A.M., Bowen, A.J., *et al.* (2013). Comprehensive Assignment of Roles for *Salmonella* Typhimurium Genes in Intestinal Colonization of Food-Producing Animals. *Plos Genetics* 9.
- Chen, X., Taylor, D.W., Fowler, C.C., Galan, J.E., Wang, H.W., and Wolin, S.L. (2013). An RNA degradation machine sculpted by Ro autoantigen and noncoding RNA. *Cell* 153, 166-177.
- Chen, Z.W., Hsuan, S.L., Liao, J.W., Chen, T.H., Wu, C.M., Lee, W.C., Lin, C.C., Liao, C.M., Yeh, K.S., Winton, J.R., *et al.* (2010). Mutations in the *Salmonella enterica* serovar *Choleraesuis* cAMP-receptor protein gene lead to functional defects in the SPI-1 Type III secretion system. *Veterinary Research* 41.
- Chertow, D.S., and Memoli, M.J. (2013). Bacterial Coinfection in Influenza A Grand Rounds Review. *Jama-J Am Med Assoc* 309, 275-282.
- Choi, J., and Groisman, E.A. (2013). The lipopolysaccharide modification regulator PmrA limits *Salmonella* virulence by repressing the type three-secretion system Spi/Ssa. *Proc Natl Acad Sci U S A* 110, 9499-9504.
- Churchman, L.S., and Weissman, J.S. (2011). Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* 469, 368-+.
- Clark, M.B., Johnston, R.L., Inostroza-Ponta, M., Fox, A.H., Fortini, E., Moscato, P., Dinger, M.E., and Mattick, J.S. (2012). Genome-wide analysis of long noncoding RNA stability. *Genome Research* 22, 885-898.
- Claudi, B., Sprote, P., Chirkova, A., Personnic, N., Zankl, J., Schurmann, N., Schmidt, A., and Bumann, D. (2014). Phenotypic variation of *salmonella* in host tissues delays eradication by antimicrobial chemotherapy. *Cell* 158, 722-733.
- Conway, T., Creecy, J.P., Maddox, S.M., Grissom, J.E., Conkle, T.L., Shadid, T.M., Teramoto, J., San Miguel, P., Shimada, T., Ishihama, A., *et al.* (2014). Unprecedented high-resolution view of bacterial operon architecture revealed by RNA sequencing. *MBio* 5, e01442-01414.
- Coombes, B.K., Wickham, M.E., Lowden, M.J., Brown, N.F., and Finlay, B.B. (2005). Negative regulation of *Salmonella* pathogenicity island 2 is required for contextual control of virulence during typhoid. *Proc Natl Acad Sci U S A* 102, 17460-17465.
- Cox, M.L., Schray, C.L., Luster, C.N., Stewart, Z.S., Korytko, P.J., KN, M.K., Paulauskis, J.D., and Dunstan, R.W. (2006). Assessment of fixatives, fixation, and tissue processing on morphology and RNA integrity. *Exp Mol Pathol* 80, 183-191.

- Croucher, N.J., and Thomson, N.R. (2010). Studying bacterial transcriptomes using RNA-seq. *Curr Opin Microbiol* 13, 619-624.
- Datsenko, K.A., and Wanner, B.L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A* 97, 6640-6645.
- de Jong, H.K., Parry, C.M., van der Poll, T., and Wiersinga, W.J. (2012). Host-pathogen interaction in invasive salmonellosis. *PLoS Pathog* 8, e1002933.
- Der, S.D., Zhou, A., Williams, B.R., and Silverman, R.H. (1998). Identification of genes differentially regulated by interferon alpha, beta, or gamma using oligonucleotide arrays. *Proc Natl Acad Sci U S A* 95, 15623-15628.
- Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel, A., Knowles, D.G., *et al.* (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res* 22, 1775-1789.
- Doose, G., Alexis, M., Kirsch, R., Findeiss, S., Langenberger, D., Machne, R., Morl, M., Hoffmann, S., and Stadler, P.F. (2013). Mapping the RNA-Seq trash bin Unusual transcripts in prokaryotic transcriptome sequencing data. *Rna Biology* 10, 1204-1210.
- Dreyfus, M., and Regnier, P. (2002). The poly(A) tail of mRNAs: Bodyguard in eukaryotes, scavenger in bacteria. *Cell* 111, 611-613.
- Dunham, I., Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C.A., Doyle, F., Epstein, C.B., Fietze, S., Harrow, J., Kaul, R., *et al.* (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57-74.
- Ellermeier, C.D., and Slauch, J.M. (2003). RtsA and RtsB coordinately regulate expression of the invasion and flagellar genes in *Salmonella enterica* serovar Typhimurium. *J Bacteriol* 185, 5096-5108.
- Ellermeier, J.R., and Slauch, J.M. (2007). Adaptation to the host environment: regulation of the SPI1 type III secretion system in *Salmonella enterica* serovar Typhimurium. *Curr Opin Microbiol* 10, 24-29.
- Eriksson, S., Lucchini, S., Thompson, A., Rhen, M., and Hinton, J.C. (2003). Unravelling the biology of macrophage infection by gene expression profiling of intracellular *Salmonella enterica*. *Mol Microbiol* 47, 103-118.
- Eulalio, A., Schulte, L., and Vogel, J. (2012). The mammalian microRNA response to bacterial infections. *RNA Biol* 9.
- Fabrega, A., and Vila, J. (2013). *Salmonella enterica* serovar Typhimurium skills to succeed in the host: virulence and regulation. *Clin Microbiol Rev* 26, 308-341.
- Fass, E., and Groisman, E.A. (2009). Control of *Salmonella* pathogenicity island-2 gene expression. *Curr Opin Microbiol* 12, 199-204.
- Faucher, S.P., Friedlander, G., Livny, J., Margalit, H., and Shuman, H.A. (2010). *Legionella pneumophila* 6S RNA optimizes intracellular multiplication. *P Natl Acad Sci USA* 107, 7533-7538.
- Fodor, S.P., Rava, R.P., Huang, X.C., Pease, A.C., Holmes, C.P., and Adams, C.L. (1993). Multiplexed biochemical assays with biological chips. *Nature* 364, 555-556.
- Förstner, Vogel, and Sharma (2014). READemption - A tool for the computational analysis of deep-sequencing-based transcriptome data. *Bioinformatics*.
- Galan, J.E., and Curtiss, R. (1989). Cloning and Molecular Characterization of Genes Whose Products Allow *Salmonella*-Typhimurium to Penetrate Tissue-Culture Cells. *P Natl Acad Sci USA* 86, 6383-6387.
- Garcia-del Portillo, F., Zwick, M.B., Leung, K.Y., and Finlay, B.B. (1993). *Salmonella* Induces the Formation of Filamentous Structures Containing Lysosomal Membrane-Glycoproteins in Epithelial-Cells. *P Natl Acad Sci USA* 90, 10544-10548.
- Georg, J., and Hess, W.R. (2011). cis-Antisense RNA, Another Level of Gene Regulation in Bacteria. *Microbiol Mol Biol R* 75, 286-+.
- Gerdes, K., and Maisonneuve, E. (2012). Bacterial Persistence and Toxin-Antitoxin Loci. *Annual Review of Microbiology*, Vol 66 66, 103-123.

- Gerlach, R.G., Jackel, D., Geymeier, N., and Hensel, M. (2007). Salmonella pathogenicity island 4-mediated adhesion is coregulated with invasion genes in *Salmonella enterica*. *Infection and Immunity* 75, 4697-4709.
- Giangrossi, M., Prosseda, G., Tran, C.N., Brandi, A., Colonna, B., and Falconi, M. (2010). A novel antisense RNA regulates at transcriptional level the virulence gene *icsA* of *Shigella flexneri*. *Nucleic Acids Research* 38, 3362-3375.
- Gillespie, J.W., Best, C.J., Bichsel, V.E., Cole, K.A., Greenhut, S.F., Hewitt, S.M., Ahram, M., Gathright, Y.B., Merino, M.J., Strausberg, R.L., *et al.* (2002). Evaluation of non-formalin tissue fixation for molecular profiling studies. *Am J Pathol* 160, 449-457.
- Gomez, J.A., Wapinski, O.L., Yang, Y.W., Bureau, J.F., Gopinath, S., Monack, D.M., Chang, H.Y., Brahic, M., and Kirkegaard, K. (2013). The NeST Long ncRNA Controls Microbial Susceptibility and Epigenetic Activation of the Interferon-gamma Locus. *Cell* 152, 743-754.
- Gong, H., Vu, G.P., Bai, Y., Chan, E., Wu, R.B., Yang, E., Liu, F.Y., and Lu, S.W. (2011). A Salmonella Small Non-Coding RNA Facilitates Bacterial Invasion and Intracellular Replication by Modulating the Expression of Virulence Factors. *Plos Pathogens* 7.
- Gonzalo-Asensio, J., Ortega, A.D., Rico-Perez, G., Pucciarelli, M.G., and Garcia-del Portillo, F. (2013). A Novel Antisense RNA from the Salmonella Virulence Plasmid pSLT Expressed by Non-Growing Bacteria inside Eukaryotic Cells. *Plos One* 8.
- Gottesman, S., and Storz, G. (2011). Bacterial Small RNA Regulators: Versatile Roles and Rapidly Evolving Variations. *Cold Spring Harbor perspectives in biology* 3.
- Gough, D.J., Corlett, A., Schlessinger, K., Wegrzyn, J., Lerner, A.C., and Levy, D.E. (2009). Mitochondrial STAT3 Supports Ras-Dependent Oncogenic Transformation. *Science* 324, 1713-1716.
- Grieshaber, N.A., Grieshaber, S.S., Fischer, E.R., and Hackstadt, T. (2006). A small RNA inhibits translation of the histone-like protein Hc1 in *Chlamydia trachomatis*. *Molecular Microbiology* 59, 541-550.
- Gripenland, J., Netterling, S., Loh, E., Tiensuu, T., Toledo-Arana, A., and Johansson, J. (2010). RNAs: regulators of bacterial virulence. *Nature Reviews Microbiology* 8, 857-866.
- Guell, M., Yus, E., Lluch-Senar, M., and Serrano, L. (2011). Bacterial transcriptomics: what is beyond the RNA hori-z-ome? *Nat Rev Microbiol* 9, 658-669.
- Guillier, M., and Gottesman, S. (2008). The 5' end of two redundant sRNAs is involved in the regulation of multiple targets, including their own regulator. *Nucleic Acids Research* 36, 6781-6794.
- Guo, L., Lim, K.B., Gunn, J.S., Bainbridge, B., Darveau, R.P., Hackett, M., and Miller, S.I. (1997). Regulation of lipid modifications by *Salmonella typhimurium* virulence genes *phoP-phoQ*. *Science* 276, 250-253.
- Guo, M.S., Updegrave, T.B., Gogol, E.B., Shabalina, S.A., Gross, C.A., and Storz, G. (2014). MicL, a new sigmaE-dependent sRNA, combats envelope stress by repressing synthesis of Lpp, the major outer membrane lipoprotein. *Genes Dev* 28, 1620-1634.
- Haas, B.J., Chin, M., Nusbaum, C., Birren, B.W., and Livny, J. (2013). How deep is deep enough for RNA-Seq profiling of bacterial transcriptomes? *BMC Genomics* 13, 734.
- Hamatani, K., Eguchi, H., Takahashi, K., Koyama, K., Mukai, M., Ito, R., Taga, M., Yasui, W., and Nakachi, K. (2006). Improved RT-PCR amplification for molecular analyses with long-term preserved formalin-fixed, paraffin-embedded tissue specimens. *J Histochem Cytochem* 54, 773-780.
- Hannemann, S., Gao, B., and Galan, J.E. (2013). Salmonella modulation of host cell gene expression promotes its intracellular growth. *PLoS Pathog* 9, e1003668.
- Haraga, A., Ohlson, M.B., and Miller, S.I. (2008). Salmonellae interplay with host cells. *Nat Rev Microbiol* 6, 53-66.
- Hardt, W.D., and Galan, J.E. (1997). A secreted Salmonella protein with homology to an avirulence determinant of plant pathogenic bacteria. *Proc Natl Acad Sci U S A* 94, 9887-9892.

- Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S., *et al.* (2012). GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res* 22, 1760-1774.
- Hautefort, I., Thompson, A., Eriksson-Ygberg, S., Parker, M.L., Lucchini, S., Danino, V., Bongaerts, R.J., Ahmad, N., Rhen, M., and Hinton, J.C. (2008). During infection of epithelial cells *Salmonella enterica* serovar Typhimurium undergoes a time-dependent transcriptional adaptation that results in simultaneous expression of three type 3 secretion systems. *Cell Microbiol* 10, 958-984.
- He, L., and Hannon, G.J. (2004). MicroRNAs: Small RNAs with a big role in gene regulation (vol 5, pg 522 2004). *Nature Reviews Genetics* 5, 522-+.
- Hebrard, M., Kroger, C., Srikumar, S., Colgan, A., Handler, K., and Hinton, J.C.D. (2012). sRNAs and the virulence of *Salmonella enterica* serovar Typhimurium. *Rna Biology* 9, 437-445.
- Hedges, J.F., Lubick, K.J., and Jutila, M.A. (2005). Gamma delta T cells respond directly to pathogen-associated molecular patterns. *J Immunol* 174, 6045-6053.
- Heidrich, N., and Vogel, J. (2013). CRISPRs extending their reach: prokaryotic RNAi protein Cas9 recruited for gene regulation. *Embo Journal* 32, 1802-1804.
- Helaine, S., Cheverton, A.M., Watson, K.G., Faure, L.M., Matthews, S.A., and Holden, D.W. (2014). Internalization of *Salmonella* by macrophages induces formation of nonreplicating persisters. *Science* 343, 204-208.
- Helaine, S., and Kugelberg, E. (2014). Bacterial persisters: formation, eradication, and experimental systems. *Trends in microbiology* 22, 417-424.
- Himeno, H., Kurita, D., and Muto, A. (2014). tmRNA-mediated trans-translation as the major ribosome rescue system in a bacterial cell. *Frontiers in genetics* 5, 66.
- Hinton, J.C., Hautefort, I., Eriksson, S., Thompson, A., and Rhen, M. (2004). Benefits and pitfalls of using microarrays to monitor bacterial gene expression during infection. *Curr Opin Microbiol* 7, 277-282.
- Hobbs, E.C., Astarita, J.L., and Storz, G. (2010). Small RNAs and small proteins involved in resistance to cell envelope stress and acid shock in *Escherichia coli*: analysis of a bar-coded mutant collection. *J Bacteriol* 192, 59-67.
- Hoe, C.H., Raabe, C.A., Rozhdestvensky, T.S., and Tang, T.H. (2013). Bacterial sRNAs: Regulation in stress. *International Journal of Medical Microbiology* 303, 217-229.
- Hoffmann, S., Otto, C., Doose, G., Tanzer, A., Langenberger, D., Christ, S., Kunz, M., Holdt, L., Teupser, D., Hackermueller, J., *et al.* (2014). A multi-split mapping algorithm for circular RNA, splicing, trans-splicing, and fusion detection. *Genome Biol* 15, R34.
- Holmqvist, E., Reimegard, J., Sterk, M., Grantcharova, N., Romling, U., and Wagner, E.G. (2010). Two antisense RNAs target the transcriptional regulator CsgD to inhibit curli synthesis. *EMBO J* 29, 1840-1850.
- Huarte, M., Guttman, M., Feldser, D., Garber, M., Koziol, M.J., Kenzelmann-Broz, D., Khalil, A.M., Zuk, O., Amit, I., Rabani, M., *et al.* (2010). A Large Intergenic Noncoding RNA Induced by p53 Mediates Global Gene Repression in the p53 Response. *Cell* 142, 409-419.
- Humphrys, M.S., Creasy, T., Sun, Y., Shetty, A.C., Chibucos, M.C., Drabek, E.F., Fraser, C.M., Farooq, U., Sengamalay, N., Ott, S., *et al.* (2013). Simultaneous transcriptional profiling of bacteria and their host cells. *PLoS One* 8, e80597.
- Hung, T., Wang, Y.L., Lin, M.F., Koegel, A.K., Kotake, Y., Grant, G.D., Horlings, H.M., Shah, N., Umbricht, C., Wang, P., *et al.* (2011). Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. *Nature Genetics* 43, 621-U196.
- Iiott, N.E., Heward, J.A., Roux, B., Tsitsiou, E., Fenwick, P.S., Lenzi, L., Goodhead, I., Hertz-Fowler, C., Heger, A., Hall, N., *et al.* (2014). Long non-coding RNAs and enhancer RNAs regulate the lipopolysaccharide-induced inflammatory response in human monocytes. *Nature Communications* 5.
- Imamura, K., Imamachi, N., Akizuki, G., Kumakura, M., Kawaguchi, A., Nagata, K., Kato, A., Kawaguchi, Y., Sato, H., Yoneda, M., *et al.* (2014). Long Noncoding RNA NEAT1-Dependent SFPQ Relocation from Promoter Region to Paraspeckle Mediates IL8 Expression upon Immune Stimuli (vol 53, pg 393, 2014). *Molecular Cell* 54, 1055-1055.

- Ingolia, N.T., Ghaemmaghami, S., Newman, J.R.S., and Weissman, J.S. (2009). Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. *Science* 324, 218-223.
- Jahn, C.E., Charkowski, A.O., and Willis, D.K. (2008). Evaluation of isolation methods and RNA integrity for bacterial RNA quantitation. *J Microbiol Meth* 75, 318-324.
- Jeck, W.R., and Sharpless, N.E. (2014). Detecting and characterizing circular RNAs. *Nature Biotechnology* 32, 453-461.
- Jeffreys, A.G., Hak, K.M., Steffan, R.J., Foster, J.W., and Bej, A.K. (1998). Growth, survival and characterization of *cspA* in *Salmonella enteritidis* following cold shock. *Current Microbiology* 36, 29-35.
- Jenner, R.G., and Young, R.A. (2005). Insights into host responses against pathogens from transcriptional profiling. *Nat Rev Microbiol* 3, 281-294.
- Jiang, L., Schlesinger, F., Davis, C.A., Zhang, Y., Li, R., Salit, M., Gingeras, T.R., and Oliver, B. (2011). Synthetic spike-in standards for RNA-seq experiments. *Genome Res* 21, 1543-1551.
- Jorgensen, M.G., Thomason, M.K., Havelund, J., Valentin-Hansen, P., and Storz, G. (2013). Dual function of the *McaS* small RNA in controlling biofilm formation. *Genes Dev* 27, 1132-1145.
- Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2014). Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res* 42, D199-205.
- Kawahara, Y., Oono, Y., Kanamori, H., Matsumoto, T., Itoh, T., and Minami, E. (2012). Simultaneous RNA-Seq Analysis of a Mixed Transcriptome of Rice and Blast Fungus Interaction. *PLoS One* 7, e49423.
- Kerns, H.M., Jutila, M.A., and Hedges, J.F. (2009). The distinct response of γ delta T cells to the Nod2 agonist muramyl dipeptide. *Cell Immunol* 257, 38-43.
- Kertesz, M., Wan, Y., Mazor, E., Rinn, J.L., Nutter, R.C., Chang, H.Y., and Segal, E. (2010). Genome-wide measurement of RNA secondary structure in yeast. *Nature* 467, 103-107.
- Kihlstrom, E. (1977). Infection of HeLa cells with *Salmonella typhimurium* 395 MS and MR10 bacteria. *Infect Immun* 17, 290-295.
- Kissinger, P., and Adamski, A. (2013). Trichomoniasis and HIV interactions: a review. *Sex Transm Infect* 89, 426-433.
- Kodzius, R., Kojima, M., Nishiyori, H., Nakamura, M., Fukuda, S., Tagami, M., Sasaki, D., Imamura, K., Kai, C., Harbers, M., *et al.* (2006). CAGE: cap analysis of gene expression. *Nature Methods* 3, 211-222.
- Kozomara, A., and Griffiths-Jones, S. (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res* 42, D68-73.
- Kroger, C., Colgan, A., Srikumar, S., Handler, K., Sivasankaran, S.K., Hammarlof, D.L., Canals, R., Grissom, J.E., Conway, T., Hokamp, K., *et al.* (2013). An infection-relevant transcriptomic compendium for *Salmonella enterica* Serovar Typhimurium. *Cell Host Microbe* 14, 683-695.
- Kroger, C., Dillon, S.C., Cameron, A.D., Papenfort, K., Sivasankaran, S.K., Hokamp, K., Chao, Y., Sittka, A., Hebrard, M., Handler, K., *et al.* (2012). The transcriptional landscape and small RNAs of *Salmonella enterica* serovar Typhimurium. *Proc Natl Acad Sci U S A* 109, E1277-1286.
- Krol, J., Loedige, I., and Filipowicz, W. (2010). The widespread regulation of microRNA biogenesis, function and decay. *Nat Rev Genet* 11, 597-610.
- Kronstad, J.W. (2006). Serial analysis of gene expression in eukaryotic pathogens. *Infect Disord Drug Targets* 6, 281-297.
- Kumar, R., Shah, P., Swiatlo, E., Burgess, S.C., Lawrence, M.L., and Nanduri, B. (2010). Identification of novel non-coding small RNAs from *Streptococcus pneumoniae* TIGR4 using high-resolution genome tiling arrays. *BMC Genomics* 11, 350.
- Kuramochi, H., Hayashi, K., Uchida, K., Miyakura, S., Shimizu, D., Vallbohmer, D., Park, S., Danenberg, K.D., Takasaki, K., and Danenberg, P.V. (2006). Vascular endothelial growth factor messenger RNA expression level is preserved in liver metastases compared with corresponding primary colorectal cancer. *Clin Cancer Res* 12, 29-33.

- Lasa, I., Toledo-Arana, A., Dobin, A., Villanueva, M., de los Mozos, I.R., Vergara-Irigaray, M., Segura, V., Fagegaltier, D., Penades, J.R., Valle, J., *et al.* (2011). Genome-wide antisense transcription drives mRNA processing in bacteria. *P Natl Acad Sci USA* *108*, 20172-20177.
- Lee, E.J., and Groisman, E.A. (2010). An antisense RNA that governs the expression kinetics of a multifunctional virulence gene. *Molecular Microbiology* *76*, 1020-1033.
- Levin, J.Z., Yassour, M., Adiconis, X., Nusbaum, C., Thompson, D.A., Friedman, N., Gnirke, A., and Regev, A. (2010). Comprehensive comparative analysis of strand-specific RNA sequencing methods. *Nat Methods* *7*, 709-715.
- Li, Y., de Haar, C., Peppelenbosch, M.P., and van der Woude, C.J. (2012a). SOCS3 in immune regulation of inflammatory bowel disease and inflammatory bowel disease-related cancer. *Cytokine & growth factor reviews* *23*, 127-138.
- Li, Z.H., Chao, T.C., Chang, K.Y., Lin, N.W., Patil, V.S., Shimizu, C., Head, S.R., Burns, J.C., and Rana, T.M. (2014). The long noncoding RNA THRIL regulates TNF alpha expression through its interaction with hnRNPL. *P Natl Acad Sci USA* *111*, 1002-1007.
- Li, Z.H., Ender, C., Meister, G., Moore, P.S., Chang, Y., and John, B. (2012b). Extensive terminal and asymmetric processing of small RNAs from rRNAs, snoRNAs, snRNAs, and tRNAs. *Nucleic Acids Research* *40*, 6787-6799.
- Lisnic, V.J., Cac, M.B., Lisnic, B., Trsan, T., Mefferd, A., Das Mukhopadhyay, C., Cook, C.H., Jonjic, S., and Trgovcich, J. (2013). Dual Analysis of the Murine Cytomegalovirus and Host Cell Transcriptomes Reveal New Aspects of the Virus-Host Cell Interface. *Plos Pathogens* *9*.
- Liu, X., Lu, R., Xia, Y., and Sun, J. (2010). Global analysis of the eukaryotic pathways and networks regulated by *Salmonella typhimurium* in mouse intestinal infection in vivo. *BMC Genomics* *11*, 722.
- Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2^{(-Delta Delta C(T))} Method. *Methods* *25*, 402-408.
- Lober, S., Jackel, D., Kaiser, N., and Hensel, M. (2006). Regulation of *Salmonella* pathogenicity island 2 genes by independent environmental signals. *International Journal of Medical Microbiology* *296*, 435-447.
- Loessner, H., Endmann, A., Leschner, S., Westphal, K., Rohde, M., Miloud, T., Hammerling, G., Neuhaus, K., and Weiss, S. (2007). Remote control of tumour-targeted *Salmonella enterica* serovar Typhimurium by the use of L-arabinose as inducer of bacterial gene expression in vivo. *Cellular Microbiology* *9*, 1529-1537.
- Loman, N.J., and Quinlan, A.R. (2014). Poretools: a toolkit for analyzing nanopore sequence data. *Bioinformatics*.
- Mandlik, A., Livny, J., Robins, W.P., Ritchie, J.M., Mekalanos, J.J., and Waldor, M.K. (2011). RNA-Seq-based monitoring of infection-linked changes in *Vibrio cholerae* gene expression. *Cell Host Microbe* *10*, 165-174.
- Mann, B., van Opijnen, T., Wang, J.M., Obert, C., Wang, Y.D., Carter, R., McGoldrick, D.J., Ridout, G., Camilli, A., Tuomanen, E.I., *et al.* (2012). Control of Virulence by Small RNAs in *Streptococcus pneumoniae*. *Plos Pathogens* *8*.
- Marioni, J.C., Mason, C.E., Mane, S.M., Stephens, M., and Gilad, Y. (2008). RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res* *18*, 1509-1517.
- Martinez, L.C., Banda, M.M., Fernandez-Mora, M., Santana, F.J., and Bustamante, V.H. (2014). HilD Induces Expression of *Salmonella* Pathogenicity Island 2 Genes by Displacing the Global Negative Regulator H-NS from *ssrAB*. *J Bacteriol* *196*, 3746-3755.
- Masse, E., and Gottesman, S. (2002). A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*. *Proc Natl Acad Sci U S A* *99*, 4620-4625.
- Masse, E., Vanderpool, C.K., and Gottesman, S. (2005). Effect of RyhB small RNA on global iron use in *Escherichia coli*. *J Bacteriol* *187*, 6962-6971.
- Matsumura, H., Reich, S., Ito, A., Saitoh, H., Kamoun, S., Winter, P., Kahl, G., Reuter, M., Kruger, D.H., and Terauchi, R. (2003). Gene expression analysis of plant host-pathogen interactions by SuperSAGE. *P Natl Acad Sci USA* *100*, 15718-15723.

- Matsuura, M., Kawasaki, K., Kawahara, K., and Mitsuyama, M. (2012). Evasion of human innate immunity without antagonizing TLR4 by mutant *Salmonella enterica* serovar Typhimurium having penta-acylated lipid A. *Innate Immun-London* 18, 764-773.
- Mattei, T.A. (2014). Biomimetic Nanopores: a New Age of DNA sequencing? *World Neurosurg.*
- Maudet, Mano, Sunkavalli, Sharan, Giacca, Förstner, and Eulalio (2014). Functional high-throughput screening identifies the miR-15 microRNA family as cellular restriction factors for *Salmonella* infection. *Nat Commun.*
- Maurer, A.P., Mehlitz, A., Mollenkopf, H.J., and Meyer, T.F. (2007). Gene expression profiles of *Chlamydomonas pneumoniae* during the developmental cycle and iron depletion-mediated persistence. *PLoS Pathog* 3, e83.
- McIntyre, L.M., Lopiano, K.K., Morse, A.M., Amin, V., Oberg, A.L., Young, L.J., and Nuzhdin, S.V. (2011). RNA-seq: technical variability and sampling. *BMC Genomics* 12, 293.
- Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., Maier, L., Mackowiak, S.D., Gregersen, L.H., Munschauer, M., *et al.* (2013). Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* 495, 333-338.
- Mercer, T.R., and Mattick, J.S. (2013). Structure and function of long noncoding RNAs in epigenetic regulation. *Nature Structural & Molecular Biology* 20, 300-307.
- Merrell, D.S., Butler, S.M., Qadri, F., Dolganov, N.A., Alam, A., Cohen, M.B., Calderwood, S.B., Schoolnik, G.K., and Camilli, A. (2002). Host-induced epidemic spread of the cholera bacterium. *Nature* 417, 642-645.
- Metzker, M.L. (2010). Sequencing technologies - the next generation. *Nat Rev Genet* 11, 31-46.
- Moon, K., Six, D.A., Lee, H.J., Raetz, C.R., and Gottesman, S. (2013). Complex transcriptional and post-transcriptional regulation of an enzyme for lipopolysaccharide modification. *Mol Microbiol* 89, 52-64.
- Morrow, B.J., Graham, J.E., and Curtiss, R. (1999). Genomic subtractive hybridization and selective capture of transcribed sequences identify a novel *Salmonella typhimurium* fimbrial operon and putative transcriptional regulator that are absent from the *Salmonella typhi* genome. *Infection and Immunity* 67, 5106-5116.
- Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 5, 621-628.
- Nakagawa, I., Nakata, M., Kawabata, S., and Hamada, S. (2004). Transcriptome analysis and gene expression profiles of early apoptosis-related genes in *Streptococcus pyogenes*-infected epithelial cells. *Cellular Microbiology* 6, 939-952.
- Naumann, M. (2000). Nuclear factor-kappa B activation and innate immune response in microbial pathogen infection. *Biochemical Pharmacology* 60, 1109-1114.
- Negrea, A., Bjur, E., Puia, S., Ygberg, S.E., Aslund, F., and Rhen, M. (2009). Thioredoxin 1 Participates in the Activity of the *Salmonella enterica* Serovar Typhimurium Pathogenicity Island 2 Type III Secretion System. *J Bacteriol* 191, 6918-6927.
- Nickels, B.E. (2012). A new way to start: nanoRNA-mediated priming of transcription initiation. *Transcription* 3, 300-304.
- Nicol, J.W., Helt, G.A., Blanchard, S.G., Raja, A., and Loraine, A.E. (2009). The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets. *Bioinformatics* 25, 2730-2731.
- Nilsen, T.W. (2008). Endo-siRNAs: yet another layer of complexity in RNA silencing. *Nature Structural & Molecular Biology* 15, 546-548.
- Oeckinghaus, A., and Ghosh, S. (2009). The NF-kappaB family of transcription factors and its regulation. *Cold Spring Harbor perspectives in biology* 1, a000034.
- Oliver, H.F., Orsi, R.H., Ponnala, L., Keich, U., Wang, W., Sun, Q., Cartinhour, S.W., Filiatrault, M.J., Wiedmann, M., and Boor, K.J. (2009). Deep RNA sequencing of *L. monocytogenes* reveals overlapping and extensive stationary phase and sigma B-dependent transcriptomes, including multiple highly transcribed noncoding RNAs. *BMC Genomics* 10, 641.
- Ortega, A.D., Gonzalo-Asensio, J., and Garcia-del Portillo, F. (2012). Dynamics of *Salmonella* small RNA expression in non-growing bacteria located inside eukaryotic cells. *RNA Biol* 9, 469-488.

- Otto, C., Stadler, P.F., and Hoffmann, S. (2014). Lacking alignments? The next-generation sequencing mapper segemehl revisited. *Bioinformatics* 30, 1837-1843.
- Paciello, I., Silipo, A., Lembo-Fazio, L., Curcuru, L., Zumsteg, A., Noel, G., Ciancarella, V., Sturiale, L., Molinaro, A., and Bernardini, M.L. (2013). Intracellular *Shigella* remodels its LPS to dampen the innate immune recognition and evade inflammasome activation. *P Natl Acad Sci USA* 110, E4345-E4354.
- Padalon-Brauch, G., Hershberg, R., Elgrably-Weiss, M., Baruch, K., Rosenshine, I., Margalit, H., and Altuvia, S. (2008). Small RNAs encoded within genetic islands of *Salmonella typhimurium* show host-induced expression and role in virulence. *Nucleic Acids Res* 36, 1913-1927.
- Papenfort, K., Pfeiffer, V., Mika, F., Lucchini, S., Hinton, J.C., and Vogel, J. (2006). SigmaE-dependent small RNAs of *Salmonella* respond to membrane stress by accelerating global omp mRNA decay. *Mol Microbiol* 62, 1674-1688.
- Papenfort, K., Podkaminski, D., Hinton, J.C., and Vogel, J. (2012). The ancestral SgrS RNA discriminates horizontally acquired *Salmonella* mRNAs through a single G-U wobble pair. *Proc Natl Acad Sci U S A* 109, E757-764.
- Papenfort, K., Said, N., Welsink, T., Lucchini, S., Hinton, J.C., and Vogel, J. (2009). Specific and pleiotropic patterns of mRNA regulation by ArcZ, a conserved, Hfq-dependent small RNA. *Mol Microbiol* 74, 139-158.
- Papenfort, K., Sun, Y., Miyakoshi, M., Vanderpool, C.K., and Vogel, J. (2013). Small RNA-Mediated Activation of Sugar Phosphatase mRNA Regulates Glucose Homeostasis. *Cell* 153, 426-437.
- Papenfort, K., and Vogel, J. (2010). Regulatory RNA in bacterial pathogens. *Cell Host Microbe* 8, 116-127.
- Papenfort, K., and Vogel, J. (2014). Small RNA functions in carbon metabolism and virulence of enteric pathogens. *Frontiers in cellular and infection microbiology* 4, 91.
- Parkhill, J., Dougan, G., James, K.D., Thomson, N.R., Pickard, D., Wain, J., Churcher, C., Mungall, K.L., Bentley, S.D., Holden, M.T., *et al.* (2001). Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. *Nature* 413, 848-852.
- Passalacqua, K.D., Varadarajan, A., Ondov, B.D., Okou, D.T., Zwick, M.E., and Bergman, N.H. (2009). Structure and complexity of a bacterial transcriptome. *J Bacteriol* 191, 3203-3211.
- Perez, N., Trevino, J., Liu, Z., Ho, S.C., Babitzke, P., and Sumby, P. (2009). A genome-wide analysis of small regulatory RNAs in the human pathogen group A *Streptococcus*. *PLoS One* 4, e7668.
- Perkins, T.T., Kingsley, R.A., Fookes, M.C., Gardner, P.P., James, K.D., Yu, L., Assefa, S.A., He, M., Croucher, N.J., Pickard, D.J., *et al.* (2009). A strand-specific RNA-Seq analysis of the transcriptome of the typhoid bacillus *Salmonella typhi*. *PLoS Genet* 5, e1000569.
- Petre, B., Morin, E., Tisserant, E., Hacquard, S., Da Silva, C., Poulain, J., Delaruelle, C., Martin, F., Rouhier, N., Kohler, A., *et al.* (2012). RNA-Seq of early-infected poplar leaves by the rust pathogen *Melampsora larici-populina* uncovers PtSultr3;5, a fungal-induced host sulfate transporter. *PLoS One* 7, e44408.
- Pfeiffer, V., Sittka, A., Tomer, R., Tedin, K., Brinkmann, V., and Vogel, J. (2007). A small non-coding RNA of the invasion gene island (SPI-1) represses outer membrane protein synthesis from the *Salmonella* core genome. *Molecular Microbiology* 66, 1174-1191.
- Prusty, B.K., Siegl, C., Hauck, P., Hain, J., Korhonen, S.J., Hiltunen-Back, E., Puolakkainen, M., Rudel, T. (2013). *Chlamydia trachomatis* infection induces replication of latent HHV-6. *Plos One* 8.
- Pusch, J., Votteler, M., Gohler, S., Engl, J., Hampel, M., Walles, H., and Schenke-Layland, K. (2011). The physiological performance of a three-dimensional model that mimics the microenvironment of the small intestine. *Biomaterials* 32, 7469-7478.
- Rapicavoli, N.A., Qu, K., Zhang, J.J., Mikhail, M., Laberge, R.M., and Chang, H.Y. (2013). A mammalian pseudogene lncRNA at the interface of inflammation and anti-inflammatory therapeutics. *Elife* 2.
- Rappuoli, R. (2000). Pushing the limits of cellular microbiology: microarrays to study bacteria-host cell intimate contacts. *Proc Natl Acad Sci U S A* 97, 13467-13469.

- Rechavi, O., Kalman, M., Fang, Y., Vernitsky, H., Jacob-Hirsch, J., Foster, L.J., Kloog, Y., and Goldstein, I. (2010). Trans-SILAC: sorting out the non-cell-autonomous proteome. *Nat Methods* 7, 923-927.
- Rehmsmeier, M., Steffen, P., Hochsmann, M., and Giegerich, R. (2004). Fast and effective prediction of microRNA/target duplexes. *Rna-a Publication of the Rna Society* 10, 1507-1517.
- Repoila, F., Majdalani, N., and Gottesman, S. (2003). Small non-coding RNAs, co-ordinators of adaptation processes in *Escherichia coli*: the RpoS paradigm. *Molecular Microbiology* 48, 855-861.
- Revel, A.T., Talaat, A.M., and Norgard, M.V. (2002). DNA microarray analysis of differential gene expression in *Borrelia burgdorferi*, the Lyme disease spirochete. *Proc Natl Acad Sci U S A* 99, 1562-1567.
- Rhodus, V.A., Suh, W.C., Nonaka, G., West, J., and Gross, C.A. (2006). Conserved and variable functions of the sigmaE stress response in related genomes. *PLoS Biol* 4, e2.
- Rinn, J.L. (2014). lncRNAs: Linking RNA to Chromatin. *Cold Spring Harbor perspectives in biology* 6.
- Rodland, K.D., Adkins, J.N., Ansong, C., Chowdhury, S., Manes, N.P., Shi, L., Yoon, H., Smith, R.D., and Heffron, F. (2008). Use of high-throughput mass spectrometry to elucidate host-pathogen interactions in *Salmonella*. *Future Microbiol* 3, 625-634.
- Romaniuk, P.J., and Uhlenbeck, O.C. (1983). Joining of Rna Molecules with Rna Ligase. *Method Enzymol* 100, 52-59.
- Romeo, T., Vakulskas, C.A., and Babitzke, P. (2013). Post-transcriptional regulation on a global scale: form and function of Csr/Rsm systems. *Environmental Microbiology* 15, 313-324.
- Romilly, C., Lays, C., Tomasini, A., Caldelari, I., Benito, Y., Hammann, P., Geissmann, T., Boisset, S., Romby, P., and Vandenesch, F. (2014). A Non-Coding RNA Promotes Bacterial Persistence and Decreases Virulence by Regulating a Regulator in *Staphylococcus aureus*. *Plos Pathogens* 10.
- Rosenstein, J. (2014). The Promise of Nanopore Technology. *Ieee Pulse* 5, 52-54.
- Rudel, T., Kepp, O., and Kozjak-Pavlovic, V. (2010). Interactions between bacterial pathogens and mitochondrial cell death pathways. *Nat Rev Microbiol* 8, 693-705.
- Saliba, A.E., Westermann, A.J., Gorski, S.A., and Vogel, J. (2014). Single-cell RNA-seq: advances and future challenges. *Nucleic Acids Res* 42, 8845-8860.
- Sampson, T.R., Saroj, S.D., Llewellyn, A.C., Tzeng, Y.L., and Weiss, D.S. (2013). A CRISPR/Cas system mediates bacterial innate immune evasion and virulence (vol 497, pg 254, 2013). *Nature* 501, 262-262.
- Santiviago, C.A., Reynolds, M.M., Porwollik, S., Choi, S.H., Long, F., Andrews-Polymenis, H.L., and McClelland, M. (2009). Analysis of pools of targeted *Salmonella* deletion mutants identifies novel genes affecting fitness during competitive infection in mice. *PLoS Pathog* 5, e1000477.
- Schena, M., Shalon, D., Davis, R.W., and Brown, P.O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270, 467-470.
- Schlumberger, M.C., and Hardt, W.D. (2006). *Salmonella* type III secretion effectors: pulling the host cell's strings. *Curr Opin Microbiol* 9, 46-54.
- Schroeder, N., Henry, T., de Chastellier, C., Zhao, W.D., Guilhon, A.A., Gorvel, J.P., and Meresse, S. (2010). The Virulence Protein SopD2 Regulates Membrane Dynamics of *Salmonella*-Containing Vacuoles. *Plos Pathogens* 6.
- Schulte, L.N., Eulalio, A., Mollenkopf, H.J., Reinhardt, R., and Vogel, J. (2011). Analysis of the host microRNA response to *Salmonella* uncovers the control of major cytokines by the let-7 family. *EMBO J* 30, 1977-1989.
- Schulte, L.N., Westermann, A.J., and Vogel, J. (2013). Differential activation and functional specialization of miR-146 and miR-155 in innate immune sensing. *Nucleic Acids Research* 41, 542-553.
- Selinger, D.W., Saxena, R.M., Cheung, K.J., Church, G.M., and Rosenow, C. (2003). Global RNA half-life analysis in *Escherichia coli* reveals positional patterns of transcript degradation. *Genome Res* 13, 216-223.

- Sesto, N., Wurtzel, O., Archambaud, C., Sorek, R., and Cossart, P. (2013). The excludon: a new concept in bacterial antisense RNA-mediated gene regulation. *Nature Reviews Microbiology* 11, 75-82.
- Sharma, C.M., Hoffmann, S., Darfeuille, F., Reignier, J., Findeiss, S., Sittka, A., Chabas, S., Reiche, K., Hackermuller, J., Reinhardt, R., *et al.* (2010). The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature* 464, 250-255.
- Sittka, A., Lucchini, S., Papenfort, K., Sharma, C.M., Rolle, K., Binnewies, T.T., Hinton, J.C., and Vogel, J. (2008). Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. *PLoS Genet* 4, e1000163.
- Sittka, A., Pfeiffer, V., Tedin, K., and Vogel, J. (2007). The RNA chaperone Hfq is essential for the virulence of *Salmonella typhimurium*. *Mol Microbiol* 63, 193-217.
- Socquet-Juglard, D., Kamber, T., Pothier, J.F., Christen, D., Gessler, C., Duffy, B., and Patocchi, A. (2013). Comparative RNA-seq analysis of early-infected peach leaves by the invasive phytopathogen *Xanthomonas arboricola* pv. *pruni*. *PLoS One* 8, e54196.
- Sorek, R., and Cossart, P. (2010). Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity. *Nat Rev Genet* 11, 9-16.
- Souquere, S., Beauclair, G., Harper, F., Fox, A., and Pierron, G. (2010). Highly Ordered Spatial Organization of the Structural Long Noncoding NEAT1 RNAs within Paraspeckle Nuclear Bodies. *Molecular Biology of the Cell* 21, 4020-4027.
- Stark, M.R., Pleiss, J.A., Deras, M., Scaringe, S.A., and Rader, S.D. (2006). An RNA ligase-mediated method for the efficient creation of large, synthetic RNAs. *Rna-a Publication of the Rna Society* 12, 2014-2019.
- Stead, M.B., Agrawal, A., Bowden, K.E., Nasir, R., Mohanty, B.K., Meagher, R.B., and Kushner, S.R. (2012). RNAsnap: a rapid, quantitative and inexpensive, method for isolating total RNA from bacteria. *Nucleic Acids Res* 40, e156.
- Sternberg, N.L., and Maurer, R. (1991). Bacteriophage-mediated generalized transduction in *Escherichia coli* and *Salmonella typhimurium*. *Methods Enzymol* 204, 18-43.
- Stewart, E.J., Aslund, F., and Beckwith, J. (1998). Disulfide bond formation in the *Escherichia coli* cytoplasm: an in vivo role reversal for the thioredoxins. *Embo Journal* 17, 5543-5550.
- Storz, G., Vogel, J., and Wassarman, K.M. (2011). Regulation by Small RNAs in Bacteria: Expanding Frontiers. *Molecular Cell* 43, 880-891.
- Sultan, M., Amstislavskiy, V., Risch, T., Schuette, M., Dokel, S., Ralser, M., Balzereit, D., Lehrach, H., and Yaspo, M.L. (2014). Influence of RNA extraction methods and library selection schemes on RNA-seq data. *BMC Genomics* 15, 675.
- Sultan, M., Schulz, M.H., Richard, H., Magen, A., Klingenhoff, A., Scherf, M., Seifert, M., Borodina, T., Soldatov, A., Parkhomchuk, D., *et al.* (2008). A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science* 321, 956-960.
- Taganov, K.D., Boldin, M.P., Chang, K.J., and Baltimore, D. (2006). NF-kappa B-dependent induction of microRNA miR-146, an inhibitor targeted to signaling proteins of innate immune responses. *P Natl Acad Sci USA* 103, 12481-12486.
- Takeshita, F., Suzuki, K., Sasaki, S., Ishii, N., Klinman, D.M., and Ishii, K.J. (2004). Transcriptional regulation of the human TLR9 gene. *Journal of Immunology* 173, 2552-2561.
- Tarca, A.L., Draghici, S., Khatri, P., Hassan, S.S., Mittal, P., Kim, J.S., Kim, C.J., Kusanovic, J.P., and Romero, R. (2009). A novel signaling pathway impact analysis. *Bioinformatics* 25, 75-82.
- The Sequencing Quality Control Consortium. (2014). A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. *Nat Biotechnol* 32, 903-914.
- Thiennimitr, P., Winter, S.E., and Baumler, A.J. (2012). *Salmonella*, the host and its microbiota. *Current Opinion in Microbiology* 15, 108-114.
- Thomason, M.K., and Storz, G. (2010). Bacterial Antisense RNAs: How Many Are There, and What Are They Doing? *Annu Rev Genet* 44, 167-188.
- Tierney, L., Linde, J., Muller, S., Brunke, S., Molina, J.C., Hube, B., Schock, U., Guthke, R., and Kuchler, K. (2012). An Interspecies Regulatory Network Inferred from Simultaneous RNA-seq of *Candida albicans* Invading Innate Immune Cells. *Front Microbiol* 3, 85.

- Toledo-Arana, A., Dussurget, O., Nikitas, G., Sesto, N., Guet-Revillet, H., Balestrino, D., Loh, E., Gripenland, J., Tiensuu, T., Vaitkevicius, K., *et al.* (2009). The *Listeria* transcriptional landscape from saprophytism to virulence. *Nature* **459**, 950-956.
- Toledo-Arana, A., Repoila, F., and Cossart, P. (2007). Small noncoding RNAs controlling pathogenesis. *Current Opinion in Microbiology* **10**, 182-188.
- Toung, J.M., Morley, M., Li, M., and Cheung, V.G. (2011). RNA-sequence analysis of human B-cells. *Genome Res* **21**, 991-998.
- Tross, D., Petrenko, L., Klaschik, S., Zhu, Q., and Klinman, D.M. (2009). Global changes in gene expression and synergistic interactions induced by TLR9 and TLR3. *Mol Immunol* **46**, 2557-2564.
- Tsolis, R.M., Baumler, A.J., Stojiljkovic, I., and Heffron, F. (1995). Fur Regulon of *Salmonella*-Typhimurium - Identification of New Iron-Regulated Genes. *J Bacteriol* **177**, 4628-4637.
- Tu, K.C., Long, T., Svenningsen, S.L., Wingreen, N.S., and Bassler, B.L. (2010). Negative Feedback Loops Involving Small Regulatory RNAs Precisely Control the *Vibrio harveyi* Quorum-Sensing Response. *Molecular Cell* **37**, 567-579.
- Tukhvatulin, A.I., Gitlin, I.I., Shcheblyakov, D.V., Artemicheva, N.M., Burdelya, L.G., Shmarov, M.M., Naroditsky, B.S., Gudkov, A.V., Gintsburg, A.L., and Logunov, D.Y. (2013). Combined stimulation of Toll-like receptor 5 and NOD1 strongly potentiates activity of NF-kappaB, resulting in enhanced innate immune reactions and resistance to *Salmonella enterica* serovar Typhimurium infection. *Infect Immun* **81**, 3855-3864.
- Uchiya, K., and Nikai, T. (2005). *Salmonella* pathogenicity island 2-dependent expression of suppressor of cytokine signaling 3 in macrophages. *Infect Immun* **73**, 5587-5594.
- Uchiya, K., and Nikai, T. (2008). *Salmonella* virulence factor SpiC is involved in expression of flagellin protein and mediates activation of the signal transduction pathways in macrophages. *Microbiology* **154**, 3491-3502.
- Udekwi, K.I., Darfeuille, F., Vogel, J., Reimegard, J., Holmqvist, E., and Wagner, E.G. (2005). Hfq-dependent regulation of OmpA synthesis is mediated by an antisense RNA. *Genes Dev* **19**, 2355-2366.
- Urban, J.H., and Vogel, J. (2007). Translational control and target recognition by *Escherichia coli* small RNAs in vivo. *Nucleic Acids Res* **35**, 1018-1037.
- van der Oost, J., Westra, E.R., Jackson, R.N., and Wiedenheft, B. (2014). Unravelling the structural and mechanistic basis of CRISPR-Cas systems. *Nature Reviews Microbiology* **12**, 479-492.
- Vannucci, F.A., Foster, D.N., and Gebhart, C.J. (2013). Laser microdissection coupled with RNA-seq analysis of porcine enterocytes infected with an obligate intracellular pathogen (*Lawsonia intracellularis*). *BMC Genomics* **14**, 421.
- Velculescu, V.E., Zhang, L., Vogelstein, B., and Kinzler, K.W. (1995). Serial Analysis of Gene-Expression. *Science* **270**, 484-487.
- Vicens, Q., and Cech, T.R. (2009). A natural ribozyme with 3',5' RNA ligase activity. *Nature Chemical Biology* **5**, 97-99.
- Vogel, J., Gottesman, S., Belasco, J., and Narberhaus, F. (2014). Regulating with RNA in Bacteria 2013. *Rna Biology* **11**, 403-412.
- Vogel, J., and Luisi, B.F. (2011). Hfq and its constellation of RNA. *Nature Reviews Microbiology* **9**, 578-589.
- Vogel, J., and Papenfort, K. (2006). Small non-coding RNAs and the bacterial outer membrane. *Current Opinion in Microbiology* **9**, 605-611.
- Wagner, E.G.H. (2013). Cycling of RNAs on Hfq. *Rna Biology* **10**, 619-626.
- Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* **10**, 57-63.
- Weinberg, Z., Perreault, J., Meyer, M.M., and Breaker, R.R. (2009). Exceptional structured noncoding RNAs revealed by bacterial metagenome analysis. *Nature* **462**, 656-U131.
- Westermann, A.J., Gorski, S.A., and Vogel, J. (2012). Dual RNA-seq of pathogen and host. *Nat Rev Microbiol* **10**, 618-630.
- Winkler, W.C., and Breaker, R.R. (2005). Regulation of bacterial gene expression by riboswitches. *Annual Review of Microbiology* **59**, 487-517.

- Winter, S.E., Raffatellu, M., Wilson, R.P., Russmann, H., and Baumler, A.J. (2008). The *Salmonella enterica* serotype Typhi regulator TviA reduces interleukin-8 production in intestinal epithelial cells by repressing flagellin secretion. *Cellular Microbiology* 10, 247-261.
- Winther, K.S., and Gerdes, K. (2011). Enteric virulence associated protein VapC inhibits translation by cleavage of initiator tRNA. *Proc Natl Acad Sci U S A* 108, 7403-7407.
- Wyllie, D.H., Kiss-Toth, E., Visintin, A., Smith, S.C., Boussof, S., Segal, D.M., Duff, G.W., and Dower, S.K. (2000). Evidence for an accessory protein function for toll-like receptor 1 in anti-bacterial responses. *Journal of Immunology* 165, 7125-7132.
- Xie, C., Yuan, J., Li, H., Li, M., Zhao, G., Bu, D., Zhu, W., Wu, W., Chen, R., and Zhao, Y. (2014). NONCODEv4: exploring the world of long non-coding RNA genes. *Nucleic Acids Res* 42, D98-103.
- Xu, M.L., and Su, Z.C. (2009). Computational prediction of cAMP receptor protein (CRP) binding sites in cyanobacterial genomes. *Bmc Genomics* 10.
- Yazawa, T., Kawahigashi, H., Matsumoto, T., and Mizuno, H. (2013). Simultaneous transcriptome analysis of *Sorghum* and *Bipolaris sorghicola* by using RNA-seq in combination with de novo transcriptome assembly. *PLoS One* 8, e62460.
- Ying, Y.L., Cao, C., and Long, Y.T. (2014). Single molecule analysis by biological nanopore sensors. *Analyst* 139, 3826-3835.
- Yoder-Himes, D.R., Chain, P.S., Zhu, Y., Wurtzel, O., Rubin, E.M., Tiedje, J.M., and Sorek, R. (2009). Mapping the *Burkholderia cenocepacia* niche response via high-throughput sequencing. *Proc Natl Acad Sci U S A* 106, 3976-3981.
- Yoon, H.J., McDermott, J.E., Porwollik, S., McClelland, M., and Heffron, F. (2009). Coordinated Regulation of Virulence during Systemic Infection of *Salmonella enterica* Serovar Typhimurium. *Plos Pathogens* 5.
- Yoshimura, A., Naka, T., and Kubo, M. (2007). SOCS proteins, cytokine signalling and immune regulation. *Nat Rev Immunol.* 7, 454-465.
- Zaitoun, I., Erickson, C.S., Schell, K., and Epstein, M.L. (2010). Use of RNAlater in fluorescence-activated cell sorting (FACS) reduces the fluorescence from GFP but not from DsRed. *BMC Res Notes* 3, 328.
- Zhao, Y., He, D., Zhao, J., Wang, L., Leff, A.R., Spannhake, E.W., Georas, S., and Natarajan, V. (2007). Lysophosphatidic acid induces interleukin-13 (IL-13) receptor alpha2 expression and inhibits IL-13 signaling in primary human bronchial epithelial cells. *J Biol Chem* 282, 10172-10179.
- Zheng, D.L., Constantinidou, C., Hobman, J.L., and Minchin, S.D. (2004). Identification of the CRP regulon using in vitro and in vivo transcriptional profiling. *Nucleic Acids Research* 32, 5874-5893.
- Zheng, X., Zheng, H., Lan, R., Ye, C., Wang, Y., Zhang, J., Jing, H., Chen, C., Segura, M., Gottschalk, M., *et al.* (2011). Identification of genes and genomic islands correlated with high pathogenicity in *Streptococcus suis* using whole genome tiling microarrays. *PLoS One* 6, e17987.
- Zhu, H., Cong, J.P., Mamtora, G., Gingeras, T., and Shenk, T. (1998). Cellular gene expression altered by human cytomegalovirus: global monitoring with oligonucleotide arrays. *Proc Natl Acad Sci U S A* 95, 14470-14475.

7. Appendix

7.1 Intracellular sRNA expression profiles in diverse host cell types

The functional insights obtained for the highly induced STnc440 sRNA in pathogen-host interaction (chapter 3) may prompt screens for the performance of additional *Salmonella* sRNA deletion mutants during infection in the future. Although an intracellular induction does not necessarily coincide with a virulence phenotype of the corresponding deletion mutant (Barquist et al., 2013), regulated sRNAs during infection are likely candidates for novel virulence-related riboregulators previously missed in mRNA-centric infection screens. Similarly to STnc440, many other sRNA loci are strongly induced upon host cell invasion (Fig. 3.5). Six of these sRNAs were selected for Northern blot validations and all six of them were induced also under SPI-2 conditions *in vitro* (Fig. 7.1). Of note, with the exception of the Y-RNAs (YrIA/B) and IsrE that were not detected in these experiments due to the lower sequencing depth, the induction of the selected sRNAs was also observed in various additional host cell types (Fig. 7.2).

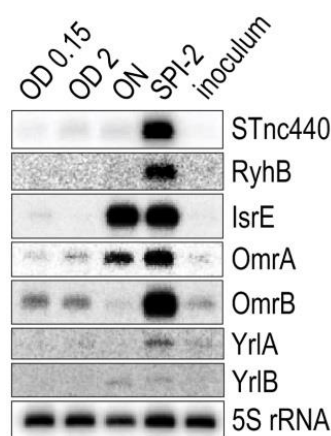


Figure 7.1|In vitro validation of intracellularly induced sRNAs. Northern blot analysis was used to confirm further *Salmonella* sRNAs that were induced upon host cell invasion (see Fig. 3.5). ‘ON’ refers to an overnight culture and ‘inoculum’ refers to bacteria grown in LB to an OD₆₀₀ of 2 and then transferred to the host cell medium DMEM.

The functionally related RyhB/IsrE are important regulators of iron homeostasis (Masse et al., 2005) and OmrA/B are involved in curli synthesis and the regulation of outer membrane proteins (Guillier and Gottesman, 2008; Holmqvist et al., 2010). These sRNAs control multiple target mRNAs by Hfq-dependent base-pairing. In contrast, YrIA and YrIB act as a scaffold for the RNA degradation machinery composed of the Ro ortholog Rsr and the exoribonuclease polynucleotide phosphorylase (PNPase), and serve an important function in rRNA maturation and decay (Chen et al., 2013). Importantly, while RyhB and IsrE have previously been suggested to play a role in *Salmonella* virulence – e.g. they were amongst the strongest induced *Salmonella*

sRNAs and counter-acted intracellular growth in a fibroblast model (Ortega et al., 2012) – OmrA/B and YrlA/B have so far not been implicated in pathogenicity. Of these sRNA genes, only *omrA* was hit in the previous transposon screen for virulence factors, but its inhibition exhibited no growth defect compared to the wild-type (Chaudhuri et al., 2013).

Pair-wise deletion of these functionally related sRNAs has been performed and the resulting mutants will subsequently be analyzed by comparative Dual RNA-seq. To increase sensitivity of the method, bacterial and host ribosomal transcripts will be removed prior to library construction using the Ribo-Zero Magnetic Gold kit (epicenter). With this strategy, even subtle changes in gene expression between mutant and wild-type infections might possibly become detectable.

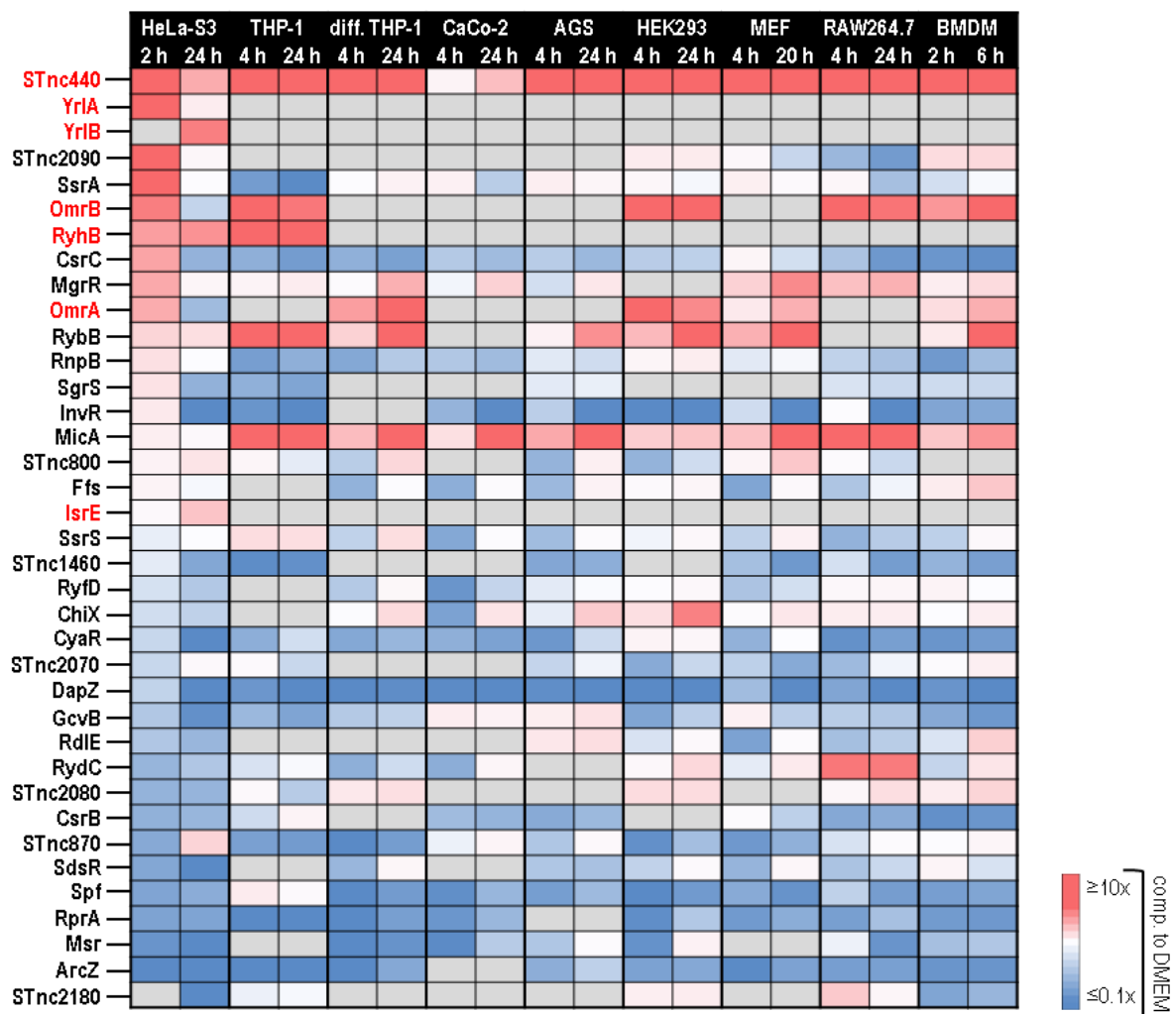


Figure 7.2/sRNA expression profile inside various host cell types. Relative expression of *Salmonella* sRNAs at an early or late time point after the infection of diverse cell types relative to their extracellular expression levels (in DMEM medium). Data are derived from a single replicate and each library was sequenced to 1-5 M reads (same as in Fig. 3.7). sRNAs selected for deletion experiments are labeled in red. Grey shaded boxes mean that the respective sRNA was not detected under this condition.

8. Curriculum vitae

Personal Details

Born: September 10th, 1984 in Bruchsal, Germany
 Nationality German

Education

October 2010 to date:
 PhD student in the Vogel Lab (RNA Biology)
 Institute of Molecular Infections Biology, Julius-Maximilian University, Würzburg, Germany
 Stipend holder of the Bavarian Elite Network (Bayrische Eliteförderung, Universität Bayern e.V.)

October 2008 – July 2010:
 Master's program in Molecular Biosciences (major: Molecular and Cellular Biology)
 Ruprecht-Karls University, Heidelberg, Germany
 Title of thesis: "Regulation of 3'End Processing and its Impact on Differentiation in a Neuroblastoma Model"

Master of Science in Molecular Biosciences: July 1st 2010

October 2005 – June 2008:
 Bachelor's program in Biology
 Ruprecht-Karls University, Heidelberg, Germany
 Title of thesis: "The Expression of Plant Polypyrimidine Tract Binding Protein Homologues Involves Alternative pre-mRNA Splicing and Is Subject to Feedback Regulation"

Bachelor of Science in Biology: June 13th 2008

September 1994 – June 2004:
 Leibniz-Gymnasium, Östringen (Academic high school)

Allgemeine Hochschulreife (University-entrance diploma) in June 2004

September 1991 – August 1994:
 Grund-und-Hauptschule Mingolsheim, Bad Schönborn (Elementary school)

Practical courses apart from/prior to the studies

Aug-Dec 2009: Four-month stipended internship as visiting scholar at the Department of Molecular and Cell Biology/Department of Chemistry, Hammond group, UC Berkeley, CA

Mar/Apr 2005: Six-week laboratory course at the University Clinic Heidelberg – Institute of Immunology (Hänsch group), Heidelberg, Germany

Feb/Mar 2005: Six-week laboratory course at the Agricultural Analysis and Research Institute Augustenberg (LUFA), Karlsruhe, Germany

Sep-Nov 2004: Three-month stay in Costa Rica, Latin America via TravelWorks (Spanish language course in San José for seven weeks, voluntary work at two national parks for two weeks each)

9. List of publications

Publications associated to the present work and beyond

The listed publications arose during the PhD work and co-operations therein:

- **Westermann AJ**, Gorski SA, Vogel J. (2012) *Dual RNA-seq of pathogen and host*. Nat Rev Microbiol. 10(9):618-30.
- Schulte LN, **Westermann AJ**, Vogel J. (2013) *Differential activation and functional specialization of miR-146 and miR-155 in innate immune sensing*. Nucleic Acids Res. 41(1):542-53.
- Saliba AE, **Westermann AJ**, Gorski SA, Vogel J. (2014) *Single-cell RNA-seq: advances and future challenges*. Nucleic Acids Res. 42(14):8845-60.

The following manuscripts related to the PhD work are in preparation or submitted:

- **Westermann AJ**, Förstner KU, Amman F, Chao Y, Schulte LN, Reinhardt R, Stadler PF, Vogel J. *Dual RNA-seq unveils noncoding RNA functions in Salmonella-host interplay*. Submitted.
- Afonso-Grunz F, Hoffmeier K, Müller S, **Westermann AJ**, Rotter B, Vogel J, Winter P, Kahl G. *Dual 3'Seq using deepSuperSAGE uncovers transcriptomes of interacting Salmonella Typhimurium and human host cells*. In preparation.

Previous publications

The listed studies were published prior to the beginning of, and are unrelated to, the PhD work:

- Stauffer E, **Westermann A**, Wagner G, Wachter A. (2010) *Polypyrimidine tract-binding protein homologues from Arabidopsis underlie regulatory circuits based on alternative splicing and downstream control*. Plant J. 64(2):243-55.
- Hickey SF, Sridhar M, **Westermann AJ**, Qin Q, Vijayendra P, Liou G, Hammond MC. (2012) *Transgene regulation in plants by alternative splicing of a suicide exon*. Nucleic Acids Res. 40(10):4701-10.

Patents

- Hammond MC, **Westermann AJ**, Qin Q. (2013) *P5sm suicide exon for regulating gene expression*. US Patent App. 13/747,395.

10. Attended conferences and courses

1. SIROCCO Small RNA Cloning Workshop
6.-9. 12. 2010
University of Cambridge, UK
2. FEMS-Leopoldina-Symposium on “Emerging topics in microbial pathogenesis”
12.-14. 4. 2011
Würzburg, Germany
Poster presentation
3. First Mol Micro Meeting
4.-6. 5. 2011
Würzburg, Germany
Poster presentation
4. NGFN-Plus Workshop “small RNAs in Infectious Diseases”
18./19. 5. 2011
Munich, Germany
5. Sixteenth Annual Meeting of the RNA Society
14.-18. 6. 2011
Kyoto, Japan
Poster presentation
6. Statusseminar “Medizinische Infektionsgenomik”
21./22. 9. 2011
Göttingen, Germany
7. RNA Summit 2012 (Mini Retreat)
23.-26. 2. 2012
Scheffau/Allgäu, Germany
8. Second Mol Micro Meeting
25.-27. 4. 2012
Würzburg, Germany
Poster presentation
9. Medizinische Infektionsgenomik: RNA-seq workshop
28./29. 6. 2012
Würzburg, Germany
Co-organizer and talk (20 min)
10. 64. Jahrestagung der Deutschen Gesellschaft für Hygiene und Mikrobiologie (DGHM) e. V.
30. 9.-3. 10. 2012
Hamburg, Germany
Poster presentation (poster award won)

11. GBM Study Group RNA-Biochemistry Meeting with Workshop 'RNA trafficking'
4.-7. 10. 2012
Bonn, Germany
Poster presentation (poster award won)
12. Cell Symposia: Functional RNAs
2.-4. 12. 2012
Sitges, Spain
Poster presentation
13. Second Course on Non-Coding Genome (Institut Curie)
10.-14. 12. 2012
Paris, France
Poster presentation and chair of a journal club session
14. Regulating with RNA in Bacteria
4.-8. 6. 2013
Würzburg, Germany
Talk (20 min); mentioned in the meeting report (Vogel et al., 2014)
15. Systems Biology of Infection Symposium
23.-27. 6. 2013
Ascona, Switzerland
Talk (15 min) and poster presentation
16. EMBO|EMBL Symposium: The Non-Coding Genome
9.-12. 10. 2013
Heidelberg, Germany
Poster presentation
17. Statusseminar "Medizinische Infektionsgenomik"
17./18. 10. 2013
Würzburg, Germany
Talk (30 min)
18. Fourth Max Planck-Genome-Centre Meeting
3. 12. 2013
Cologne, Germany
Talk (25 min)
19. EMBO Global Exchange Lecture Course "Biology of bacterial non-coding RNAs"
4.-9. 3. 2014
Buenos Aires, Argentina
Lecture on RNA-seq (15 min) and poster presentation
20. Second International Congress "Metabolism meets Virulence"
6.-9. 4. 2014
Osnabrück, Germany
Talk (20 min)

21. Third Mol Micro Meeting
7.-9. 5. 2014
Würzburg, Germany
Talk (15 min)
22. Third Young Microbiologists Symposium on Microbe Signalling, Organisation and Pathogenesis
2./3. 6. 2014
Dundee, Scotland
Talk (20 min); mentioned in the meeting report (Caly et al., 2014)
23. 7. Tagung der Fachgruppe Mikrobielle Pathogenität
16.-18. 6. 2014
Bad Urach, Germany
Talk (20 min)
24. 4. Gemeinsame Tagung der Vereinigung für Allgemeine und Angewandte Mikrobiologie (VAAM) 2014 e. V. und der Deutschen Gesellschaft für Hygiene und Mikrobiologie (DGHM) e. V.
5.-8. 10. 2014
Dresden, Germany
Poster presentation
25. EMBO Practical Course "Non-coding RNA in infection"
12.-18. 10. 2014
Würzburg, Germany
Teaching

11. Contribution by others

The work described in this doctoral thesis was conducted under the supervision of Prof. Dr. Jörg Vogel at the Institute for Molecular Infection Biology (IMIB) at the University of Würzburg, Würzburg, Germany. Several parts of the work described in this dissertation have been contributed by others and are indicated below.

- The Dual RNA-seq read mapping pipeline (see Fig. 2.9) was established by Dr. Konrad Förstner (IMIB Würzburg).
- Processing of RNA-seq data and bioinformatics analyses were mainly performed by Dr. Konrad Förstner with assistance of Dr. Lars Barquist (IMIB Würzburg). Human computational analyses were performed in collaboration with the group of Prof. Dr. Peter Stadler by Dr. Fabian Amman (University of Leipzig, Leipzig, Germany). Dr. Steve Hoffmann (University of Leipzig, Leipzig, Germany) was involved in the establishment phase of the computational analysis pipeline for Dual RNA-seq datasets.
- The functional characterization of STnc440 sRNA was performed with the great help of Dr. Yanjie Chao (IMIB Würzburg), who substantially contributed to this part of the work both intellectually and experimentally.
- Wet lab experiments were conducted with the help of the technical assistants Lisa Pfeuffer (cell culture, infection assays, etc.) and Barbara Plaschke (cloning, *Salmonella in vitro* experiments). Dr. Victoria McParland sequenced cDNA libraries on the MiSeq platform.
- Routine works (e.g. passaging of cultured cells) were mostly done by Lisa Pfeuffer.
- Several synthetic oligonucleotides (Tab. 5.7), plasmids (Tab. 5.8) and *Salmonella* strains (Tab. 5.9) used in this study were designed or constructed by other group members. Especially plasmids and strains used in the third chapter were mostly cloned by Dr. Yanjie Chao with the assistance of Barbara Plaschke.
- High-throughput sequencing was performed by Dr. Richard Reinhardt and his team at the Max-Planck Institute for Plant Breeding Research, Cologne, Germany.
- cDNA libraries were constructed by the *vertis* Biotechnologie AG in Freising-Weihenstephan, Germany (<http://www.vertis-biotech.com>).

12. Acknowledgements

I would particularly like to thank my mentor, Prof. Jörg Vogel, for giving me the opportunity to perform this work in his group, embedded in such an excellent scientific environment, for his instructive supervision and continuous guidance over the years, and for encouraging me to present my work on many national and international conferences.

I am thankful to Prof. Thomas Hünig and Prof. Martin Eilers for being part of my annual PhD committee and for their precious feedback on my work. I thank Prof. Thomas Rudel for his willingness to evaluate this thesis and Prof. Thomas Dandekar for chairing the doctoral committee.

I express my gratitude to all present and former colleagues in the Vogel lab – especially Yanjie Chao, Leon Schulte and Konrad Förstner – for their continuous help, intellectual input and the harmonic working atmosphere.

I thank Lisa Pfeuffer and Barbara Plaschke for great technical assistance over the years.

I thank the “Leipziger bioinformaticians” (Prof. Peter Stadler, Fabian Amman, Steve Hoffmann) for this fruitful collaboration.

Thanks to Hilde Merkert, Caroline Tawk and the group of Ana Eulalio for help with the confocal microscope.

I am very grateful to Stan Gorski, Antoine-Emmanuel Saliba, Leon Schulte and Yanjie Chao for their critical comments and helpful discussions on the present thesis.

I thank the Graduate School of Life Sciences (GSLs) of the University of Würzburg for giving me the possibility to become part of their outstanding, interdisciplinary program and the Bavarian Elite Network (Elitenetzwerk der Universität Bayern e.V.) for receiving their highly prestigious stipend over four years.

Thanks to my family!!

Thanks to Sandy!!!

13. Affidavit/Eidesstattliche Erklärung

I hereby confirm that my thesis entitled “*Dual RNA-seq of pathogen and host*” is the result of my own work. I did not receive any help or support from commercial consultants. All sources and/or materials applied are listed and specified in the thesis.

Furthermore, I confirm that this thesis has not yet been submitted as part of another examination process neither in identical nor in similar form.

Würzburg, December 3rd, 2014

Alexander J. Westermann

Hiermit erkläre ich an Eides statt, die Dissertation „*Duale RNA-Sequenzierung eines Pathogens und seines Wirts*“ eigenständig, d.h. insbesondere selbstständig und ohne Hilfe eines kommerziellen Promotionsberaters, angefertigt und keine anderen als die von mir angegebenen Quellen und Hilfsmittel verwendet zu haben.

Ich erkläre außerdem, dass die Dissertation weder in gleicher noch in ähnlicher Form bereits in einem anderen Prüfungsverfahren vorgelegen hat.

Würzburg, den 3.12.2014

Alexander J. Westermann